

第七届中国模式识别与计算机视觉大会

The 7th Chinese Conference on
Pattern Recognition and Computer Vision (PRCV 2024)

程 序 册

主办单位: 中国自动化学会、中国图象图形学学会
中国人工智能学会、中国计算机学会

承办单位: 新疆大学

协办单位: 新疆维吾尔自治区科学技术协会、中科视拓（南京）科技有限
公司、中国计算机学会乌鲁木齐分部

赞助单位: Auto DL、腾讯优图、合合信息、思腾合力、中科睿途、驱动云

2024年10月

中国 乌鲁木齐

目 录

温馨提示	1
大会总体日程	8
指导委员会	17
组织委员会	18
主旨报告信息及讲者简介	19
特邀报告信息及讲者简介	25
专题论坛日程	33
讲习班信息与讲者简介	121
口头报告一览表	136
墙报论文一览表	138
博士生论坛信息	165
竞赛论坛	176
城市概括	177

温馨提示

欢迎参加《第七届中国模式识别与计算机视觉大会》，我们将竭诚为您服务。请对我们的工作给予支持。

（一）参会指引

场馆地址：乌鲁木齐市水磨沟区红光山路3号

会议地点：乌鲁木齐新疆国际会展中心

会议时间：10月18日—20日

交通参考：-距地窝堡国际机场 T1、T2，21 公里，乘坐出租车约 25 分钟

-距地窝堡国际机场 T3，22 公里，乘坐出租车约 26 分钟

-距乌鲁木齐站 12 公里，乘坐出租车约 20 分钟

-距乌鲁木齐南站 16 公里，乘坐出租车约 25 分钟

公交：会展中心站

（二）餐饮须知：

10月17日

晚餐 19:00—22:30 （一楼 U CAFE 西餐厅 自助晚餐）

10月18日

茶歇 11:45—12:00

午餐 14:00—15:30 （会展中心一楼登录大厅（中2门右侧））

茶歇 17:15—17:30

晚餐 19:00—22:30 （一楼 U CAFE 西餐厅 自助晚餐）

10月19日

茶歇 12:20—12:30

午餐 14:00—15:30 （会展中心一楼登录大厅（中2门右侧））

茶歇 17:15—17:30

* 晚宴及颁奖礼

时间 20:00—22:00

地点 乌鲁木齐希尔顿酒店二楼疆山厅

10月20日

茶歇 12:23—12:40

午餐 14:00—15:30 （会展中心一楼登录大厅（中2门右侧））

茶歇 17:15—17:30

晚餐 19:00—22:30 （乌鲁木齐希尔顿酒店二楼疆山厅 自助晚餐）

（三）酒店推荐

1、乌鲁木齐希尔顿酒店（国际 5 星级）。

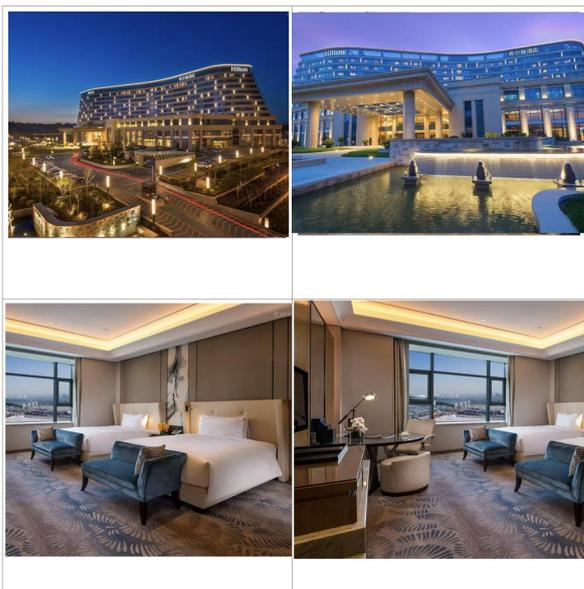
地址：新疆维吾尔自治区乌鲁木齐市米东区红光路 1237 号

房型：标间/大床

房价：650 元（含早餐）

预定电话：+86 991 783 8888

* 距离会展中心 400 米（会议自助餐地）



2、乌鲁木齐会展中心万达悦华酒店。

地址：新疆维吾尔自治区乌鲁木齐市水磨沟区红光山路 888 号留香苑

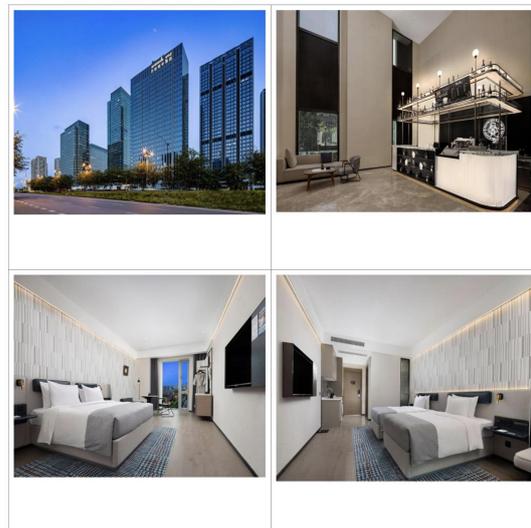
房型：标间/大床

房价：500 元（含早餐）

预定电话：18603942908

联系人：张航博

* 距离会展中心走路 1 公里



3、全季酒店（乌鲁木齐红光山会展中心酒店）。

地址：新疆维吾尔自治区乌鲁木齐市水磨沟区红光山路 888 号

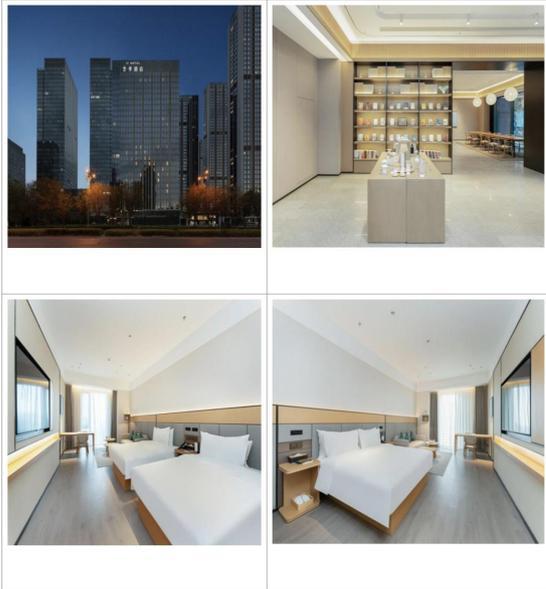
房型：标间/大床

房价：430 元（含早餐）

预定电话：0991-7785555

联系人：张苏新

* 距离会展中心走路 1 公里



4、乌鲁木齐会展中心万信至格酒店。

地址：新疆维吾尔自治区乌鲁木齐市水磨沟区会展中心会展南路 555 号万信至格酒店

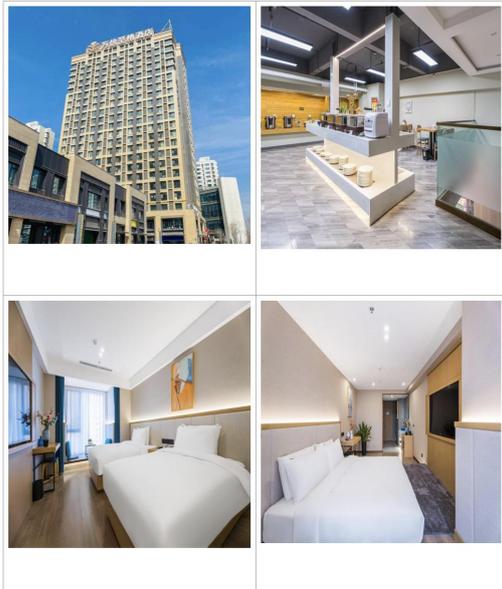
房型：标间/大床

房价：大床 400 元 标准：425 元（含早餐）

预定电话：13999268689

联系人：张钦玲

* 距离会展中心 900 米



5、麗枫酒店（乌鲁木齐红光山国际会展中心店）。已满房！！！！

地址：新疆维吾尔自治区乌鲁木齐市水磨沟区会展中心

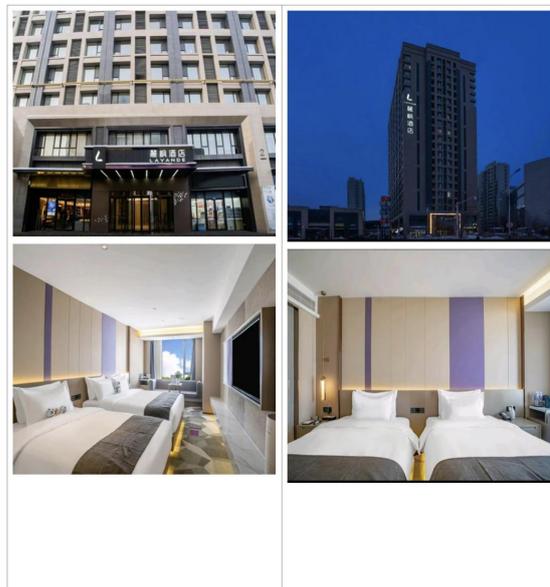
房型：标间/大床

房价：370 元（含早餐）

预定电话：17332391307

联系人：牛香

* 距离会展中心 900 米



6、乌鲁木齐中建福朋喜来登酒店。

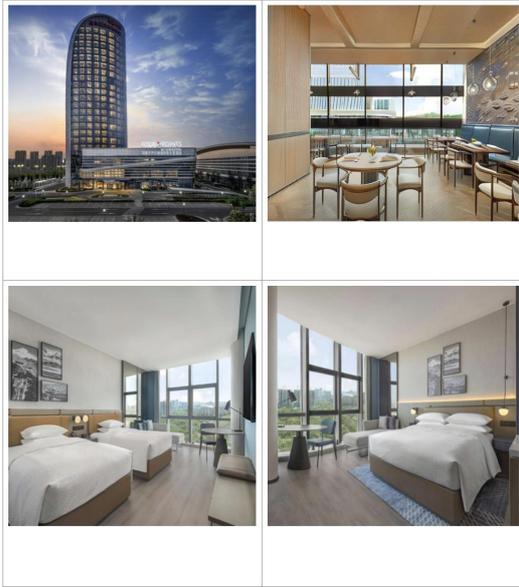
地址：新疆维吾尔自治区乌鲁木齐市米东区会展大道 4000 号

房型：标间/大床

房价：550 元（单早双早统一价）

预定电话：15199000692

* 距离会展中心 2.2 公里



(四) 天气预报

日期	天气	温度	风力
2024年10月18日	大雨	1~4℃	南风 1 级
2024年10月19日	晴天	3~8℃	东南风 1 级
2024年10月20日	小雨	1~6℃	北风 1 级

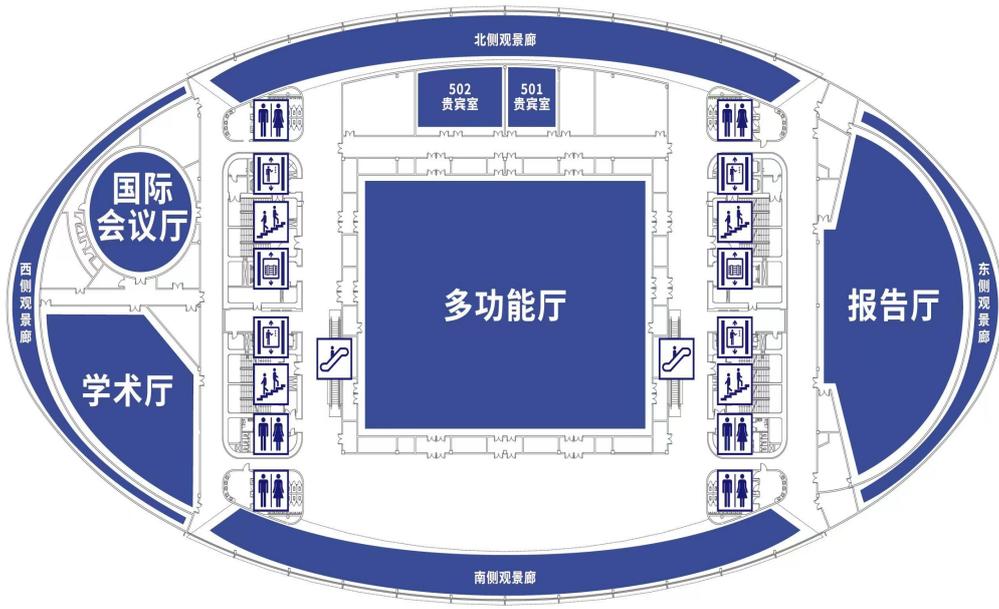
(乌鲁木齐市气象局电话：0991-2653762)

温馨提示：天气多变，外出时请您携带雨具。

(五) 会场平面图

新疆国际会展中心总平面图





五楼会议室平面图

大会总体日程

2024年10月18日 分会场：多功能厅			
14:00-15:30	午餐、午休		
15:30-17:15	讲习班：AIGC 在生物特征识别与安全中的应用	主持人：孙哲南	讲者：朱翔昱、彭勃、李琦、赵健
17:15-17:30	茶歇		
17:30-18:00	专题论坛：三维重建与生成 主持人：陶文兵、郭裕兰 讲者：申抒含、高林、彭思达、惠乐	申抒含 中国科学院自动化研究所	大规模场景高精度位姿解算与结构化三维建模
18:00-18:30		高林 中国科学院计算技术研究所	面向高真实感建模的三维重建与生成方法研究
18:30-19:00		彭思达 浙江大学	面向物理世界仿真的三维重建与生成
19:00-19:30		惠乐 西北工业大学	基于3D几何感知的可变形高斯溅射新视角合成
2024年10月18日 分会场：学术厅			
10:00-11:45	讲习班：场景感知：从生物视觉机理到类脑视觉模型	主持人：李永杰	讲者：邢大军、杨开富
11:45-12:00	茶歇		
12:00-14:00	讲习班：医疗大模型	主持人：雷柏英、胡众义	讲者：何晖光、陈浩、陈俊颖、杨二昆、房钰棋
14:00-15:30	午餐、午休		
15:30-17:15	讲习班：SfM：大规模场景三维重建的基石	主持人：申抒含、崔兆鹏、崔海楠、高翔	讲者：申抒含、崔海楠、高翔、崔兆鹏
17:15-17:30	茶歇		
17:30-17:50	专题论坛：情智兼备数字人与机器人的研究 主持人：赵思成、高跃、杨巨峰、姚鸿勋 讲者：胡斌、刘青山、刘洪海、郑文明	胡斌 北京理工大学	心智工作机制与精神障碍诊疗
17:50-18:10		刘青山 南京邮电大学	多模态视觉情感计算
18:10-18:30		刘洪海 哈尔滨工业大学（深圳）	情智兼备的孤独症早期诊疗

18:30-18:50		郑文明 东南大学	多模态情感人机交互与生理心理信息感知
18:50-19:30		所有报告嘉宾与组织者	Panel 讨论
2024年10月18日 分会场：塔里木厅			
10:00-11:45	讲习班：多模态基础模型的结构设计与优化	主持人：黄高、刘静、杨乐	讲者：周奕毅、郭龙腾、王文海
11:45-12:00	茶歇		
12:00-14:00	讲习班：基于深度学习的智能疾病诊断与个性化手术规划	主持人：陈梁骏	讲者：吴烨、杨鹏、武佳懿、马磊、张栋
14:00-15:30	午餐、午休		
15:30-15:35	女科学家论坛 主持人：贲晔、陈俊颖 讲者：张艳宁、董晶、杨欣、张铭津、王路	贲晔 山东大学 陈俊颖 华南理工大学	开场白
15:35-15:55		张艳宁 西北工业大学	以文为媒：高阶语义知识引导的视觉重建、感知与理解
15:55-16:15		董晶 中国科学院自动化研究所	数字内容生成与可信鉴伪研究浅析
16:15-16:35		杨欣 华中科技大学	视觉空间计算方法及应用
16:35-16:55		张铭津 西安电子科技大学	跨域推理赋能的智能图像处理
16:55-17:15		王路 哈尔滨工程大学	海杂波耦合情况下舰船尾迹检测技术
17:15-17:30		茶歇	
17:30-17:50	专题论坛：多模态连续学习 主持人：尹宝才、王博岳、胡永利、高源 讲者：邓成、魏云超、赵鑫、崔振、刘夏雷、高源	邓成 西安电子科技大学	在线增量学习
17:50-18:10		魏云超 北京交通大学	Continual Learning Meets Real-World Visual Perception
18:10-18:30		赵鑫 中国人民大学	面向大模型训练的数据工程方法
18:30-18:50		崔振 南京理工大学	小样本增强式持续学习研究
18:50-19:10		刘夏雷 南开大学	基于图文预训练模型的连续学习方法研究
19:10-19:30		高源 武汉大学	General-Purpose Multi-Task Learning

			with Adaptive Task-Knowledge Sharing
2024年10月18日 分会场：报告厅			
10:00-10:24	专题论坛：视觉基础模型及其前沿应用 主持人：罗斌、范登平 讲者：郑伟诗、王兴刚、叶茫、任文琦、王文海	郑伟诗 中山大学	具身视觉感知与学习
10:24-10:48		王兴刚 华中科技大学	低复杂度视觉基础模型研究
10:48-11:12		叶茫 武汉大学	多模态基础模型驱动的图文跨模态检索
11:12-11:36		任文琦 中山大学	低质量与对抗环境下的目标检测
11:36-12:00		王文海 上海人工智能实验室青年科学家	视觉语言大模型的研究与应用：通过开源套件缩小与商用多模态大模型的性能差距
12:00-14:00	CAAI-PR 专委年度会议		主持人：周杰
14:00-15:00	午餐、午休		
15:00-17:00	CSIG-DIAR 专委年度会议		主持人：金连文
17:00-19:00	CSIG-MV 专委年度会议		主持人：林宙辰
19:00-19:30	晚餐		
19:30-21:00	CAA-PRMI 专委年度会议		主持人：刘成林
21:00-23:00	CCF-CV 专委年度会议		主持人：陈熙霖
2024年10月19日 主会场：丝路印象厅			
10:00-10:30	开幕式		主持人：刘成林
10:30-11:20	主旨报告：大模型背景下的数字内容取证	报告人：谭铁牛	主持人：周杰
11:20-12:20	口头报告：最佳论文与最佳学生论文候选 1	主持人：钱育蓉、刘锦	
12:20-12:30	茶歇		
12:30-13:20	主旨报告：Understanding Egocentric Visual Attention and Actions	报告人：Yoichi Sato	主持人：查红彬
13:20-14:00	PRCV26 竞选		主持人：于仕琪
14:00-15:30	午餐、午休		
2024年10月19日 分会场：多功能厅			
15:30-16:00	特邀报告：开放条件下的视觉识别	报告人：王立君	主持人：刘夏雷
16:00-17:15	口头报告：Recognition 1		主持人：刘夏雷
17:15-17:30	茶歇		

17:30-18:00	专题论坛：机器视觉与学习 主持人：林宙辰、王奕森 讲者：尤著宏、李建欣、李霄剑、江劭玮	尤著宏 西北工业大学	生物医学知识图谱表示学习：在分子相互作用预测中的应用
18:00-18:30		李建欣 北京航空航天大学	大数据关联建模及长序预测方法
18:30-19:00		李霄剑 合肥工业大学	三维感知驱动的微创手术智能化技术研究
19:00-19:30		江劭玮 杭州电子科技大学	高通量叠层显微成像技术与应用
2024年10月19日 乌鲁木齐希尔顿酒店二楼疆山厅			
20:00-22:00	晚宴及颁奖礼		
2024年10月19日 分会场：学术厅			
15:30-16:00	特邀报告：细粒度多模态运动分析	报告人：彭宇新	主持人：刘光灿
16:00-17:15	口头报告：Recognition 2	主持人：刘光灿	
17:15-17:30	茶歇		
17:30-18:00	专题论坛：开放环境可信感知模型和学习 主持人：刘成林、胡清华、朱飞 讲者：龚怡宏、朱军、程明明、张长青、张煦尧	龚怡宏 西安交通大学	受脑认知启发的小样本持续学习方法
18:00-18:30		朱军 清华大学	从生成模型的角度探索高效鲁棒的模式识别方法
18:30-19:00		程明明 南开大学	开放动态场景下新类别发现与更新
19:00-19:20		张长青 天津大学	低质多模态数据融合
19:20-19:40		张煦尧 中科院自动化所	模式识别置信度估计
20:00-22:00	晚宴及颁奖礼		
2024年10月19日 分会场：塔里木厅			
10:30-12:20	Visual Intelligence 编委会议	主持人：查红彬	
12:20-12:30	茶歇		
12:30-14:00	PRCV 2025 筹备会议	主持人：熊红凯	
14:00-15:30	午餐、午休		
15:30-16:00	特邀报告：文本提示“检测分割一切”模型研究	报告人：王兴刚	主持人：潘金山
16:00-17:15	口头报告：Segmentation	主持人：潘金山	
17:15-17:30	茶歇		
17:30-17:55	专题论坛：可信具身视觉特征学习	吴小俊 江南大学	黎曼流形上的深度学习
17:55-18:20	主持人：孙仕亮、赵才	孟德宇	机器学习的数学技

	荣、朱磊、叶茫、殷俊	西安交通大学	术
18:20-18:45	讲者：吴小俊、孟德宇、严严、邹征夏、徐婧林	严严 厦门大学	高精度人脸表情识别技术研究
18:45-19:10		邹征夏 北京航空航天大学	生成式遥感基础模型初探
19:10-19:35		徐婧林 北京科技大学	面向行为理解的细粒度运动分析
20:00-22:00	晚宴及颁奖礼		
2024年10月19日 分会场：报告厅			
12:00-13:30	CSIG-BVD 专委年度会议	主持人：郑伟诗	
14:00-15:30	午休、午餐		
15:30-16:00	特邀报告：视觉物体检测的新趋势和新范式	报告人：张兆翔	主持人：刘日升
16:00-17:15	口头报告：Detection	主持人：刘日升	
17:15-17:30	茶歇		
17:30-17:50	专题论坛：视言碰撞：语言模型与视觉生态协同 主持人：程明明、王亚星、贾旭 讲者：徐畅、任博、盛律、王毅、杨灵	程明明 南开大学	开场白
17:50-18:10		徐畅 悉尼大学	Pixels as Parameters in Diffusion Models
18:10-18:30		任博 南开大学	含复杂光路场景的三维神经渲染
18:30-18:50		盛律 北京航空航天大学	Multimodal LLMs Meet Embodied Agents
18:50-19:10		王毅 上海人工智能实验室	InternVideo2: Scaling Video Foundation Models for Multimodal Video Understanding
19:10-19:30	杨灵 北京大学	扩散模型的前沿算法与应用	
20:00-22:00	晚宴及颁奖礼		
2024年10月19日 走廊			
11:20-12:30	海报 1		
17:15-18:25	海报 2		
2024年10月20日 主会场：丝路印象厅			
10:00-10:50	主旨报告：大模型驱动的具身智能人形机器人与展望	报告人：王耀南	主持人：林宙辰
10:50-11:21	赞助报告：AutoDL（15分钟）、腾讯优图（8分钟）、上海合合信息科技股份有限公司（8分钟）	主持人：殷绪成	
11:22-12:22	口头报告：最佳论文与最佳学生	主持人：夏威、阿力木江·艾沙	

	论文候选 2		
12:23-12:40	茶歇		
12:40-13:30	主旨报告：透明且一致的深度表征学习：从黑盒到白盒，从开环到闭环	报告人：马毅	主持人：库尔班·吾布力
13:30-13:42	赞助报告：北京思腾合力科技有限公司（6分钟）、北京中科睿途科技有限公司（3分钟）、北京趋动科技有限公司（3分钟）	主持人：吕小毅	
13:42-14:00	最佳论文与最佳学生论文颁奖	主持人：程明明	
14:00-15:30	午餐、午休		
2024年10月20日 分会场：多功能厅			
15:30-16:00	特邀报告：基于能力定量表征的序列图像自主导航技术	报告人：王大轶	主持人：艾斯卡尔·艾木都拉
16:00-17:15	口头报告：Low-level Vision	主持人：任健康	
17:15-17:30	茶歇		
17:30-18:00	专题论坛：遥感图像智能解译与应用 主持人：贺威、董燕妮、张磊、高常鑫、程堃、李淼 讲者：吴立刚、高连如、李盛阳、杨文	吴立刚 哈尔滨工业大学	空间非合作目标的智能感知与控制
18:00-18:30		高连如 中国科学院空天信息创新研究院	高光谱遥感图像目标探测技术
18:30-19:00		李盛阳 中国科学院大学	空间对地观测视频智能处理-挑战、数据集与方法
19:00-19:30		杨文 武汉大学	高分光学遥感图像小目标检测
2024年10月20日 分会场：学术厅			
15:30-16:00	特邀报告：多模态大模型发展及垂域赋能	报告人：杜博	主持人：张长青
16:00-17:15	口头报告：Applications	主持人：张长青	
17:15-17:30	茶歇		
17:30-19:30	竞赛论坛	主持人：桑农、刘偲	
2024年10月20日 分会场：塔里木厅			
15:30-16:00	特邀报告：通用数据预训练与高效微调方法研究	报告人：常建龙	主持人：王博岳
16:00-17:15	口头报告：Learning	主持人：王博岳	
17:15-17:30	茶歇		
17:30-17:40	博士生论坛 主持人：董晶、左旺孟	董晶，中国科学院自动化研究所	开场介绍
17:40-20:30	交流嘉宾：胡清华、程明明、魏云超、高跃、	常东良，北京邮电大学	每人7分钟报告，3分钟交流

	<p>姚鸿勋、邓成 博士生：常东良、陈翔、 段沛奇、顾峥、黄文柯、 寇人可、李煜东、李泽 宇、梁科、刘猛、马梅、 马彦彪、汪茜茜、伍风 翼、杨杰、周尚辰、朱 明丽</p>	<p>报告题目：基于信 息融合的细粒度图 像分类方法研究</p> <p>陈翔，南京理工大 学 报告题目：恶劣环 境下的视觉增强与 计算</p> <p>段沛奇，北京大 学 报告题目：神经形 态事件与传统图像 的互补增强机制</p> <p>顾峥，南京大 学 报告题目：基于扩 散模型的视觉上下 文学习</p> <p>黄文柯，武汉大 学 报告题目：异构联 邦学习</p> <p>寇人可，陆军工 程 大学石家庄校区 报告题目：红外小 目标检测技术进展 与开源测评框架探 索</p> <p>李煜东，深圳大 学 报告题目：人脸图 文多模态生成式预 训练模型</p> <p>李泽宇，北京邮 电 大学 报告题目：基于 2D 先验的 3D 资产材 质生成</p> <p>梁科，国防科技大 学 报告题目：结构化</p>	
--	--	---	--

		<p>知识表示与推理</p> <p>刘猛，国防科技大学 报告题目：深度时序图聚类</p> <p>马梅，宁夏大学 报告题目：基于图第一性原理的模式生成</p> <p>马彦彪 西安电子科技大学 报告题目：通用场景下深度神经网络不公平问题的几何机制与优化</p> <p>汪茜茜，安徽大学 报告题目：基于结构化 Transformer 模型的视觉表示学习</p> <p>伍风翼，电子科技大学 报告题目：基于模型驱动的红外小目标检测网络</p> <p>杨杰，香港中文大学深圳 报告题目：以人为中心的感知，理解，推理和生成</p> <p>周尚辰，新加坡南洋理工大学 报告题目： Unleashing Diffusion Models for Low-Level Vision</p> <p>朱明丽，香港中文</p>	
--	--	--	--

		大学（深圳） 报告题目： Breaking the False Sense of Security in Backdoor Defense through Re-Activation Attack	
20:30-21:00	胡清华（天津大学）， 程明明（南开大学）， 魏云超（北京交通大学）， 高跃（清华大学）， 姚鸿勋（哈尔滨工业大学）， 邓成（西安电子科技大学）	与科研专家面对面交流	
2024年10月20日 分会场：报告厅			
15:30-16:00	特邀报告：高真实感 3D 高斯数字化身	报告人：刘烨斌	主持人：郭裕兰
16:00-17:15	口头报告：3D Vision	主持人：郭裕兰	
17:15-17:30	茶歇		
17:30-17:55	专题论坛：大模型赋能 智慧医疗 主持人：陈浩、庄吓海、 徐军 讲者：王满宁、王连生、 张永兵、余晋刚、雷柏英	王满宁 复旦大学	基于大模型的病理 图像小样本学习
17:55-18:20		王连生 厦门大学	病理数据的多模态 分析
18:20-18:45		张永兵 哈尔滨工业大学（深圳）	全切片病理扫描成 像与计算
18:45-19:10		余晋刚 华南理工大学	迈向临床实用级组 织病理学 AI 辅助诊 断
19:10-19:35		雷柏英 深圳大学	大模型赋能阿尔茨 海默病智能诊断
2024年10月20日 走廊			
12:30-13:40	海报 3		
17:15-18:25	海报 4		

指导委员会

主席： 谭铁牛（中国科学院自动化研究所）

委员：（按姓氏拼音排序）

陈熙霖（中国科学院计算技术研究所）

刘成林（中国科学院自动化研究所）

芮 勇（联想集团）

查红彬（北京大学）

郑南宁（西安交通大学）

周 杰（清华大学）

秘书长： 王 亮（中国科学院自动化研究所）

组织委员会

大会主席

吾守尔·斯拉木(新疆大学)
查红彬(北京大学)

刘成林(中国科学院自动化所)
周杰(清华大学)

程序委员会主席

林宙辰(北京大学)
赫然(中国科学院自动化所)

程明明(南开大学)
库尔班·吾布力(新疆大学)

组委会主席

白翔(华中科技大学)
钱育蓉(新疆大学)

马占宇(北京邮电大学)
阿力木江·艾沙(新疆大学)

宣传主席

金连文(华南理工大学)

赵才荣(同济大学)

国际联络主席

陈熙霖(中国科学院计算所)

鲁继文(清华大学)

讲习班主席

韩军伟(西北工业大学)
雷柏英(深圳大学)

张兆翔(中国科学院自动化所)

专题论坛主席

夏桂松(武汉大学)
张姗姗(南京理工大学)

张召(合肥工业大学)

博士生论坛主席

左旺孟(哈尔滨工业大学)

董晶(中国科学院自动化所)

出版主席

张煦尧(中国科学院自动化所)

吕小毅(新疆大学)

赞助主席

殷绪成(北京科技大学)
沈超(西安交通大学)

贾伟(合肥工业大学)

展示主席

王琦(西北工业大学)
王佳佳(新疆大学)

樊彬(北京科技大学)

竞赛主席

桑农(华中科技大学)

刘偲(北京航空航天大学)

网站主席

努尔麦麦提·尤鲁瓦斯(新疆大学)

注册主席

徐学斌(新疆大学)

主旨报告信息及讲者简介

大会报告（一）



报告人：谭铁牛(南京大学)

报告题目：大模型背景下的数字内容取证

报告摘要：在数字化时代，随着人工智能技术的飞速发展，特别是深度学习技术的广泛应用，数字内容的生成和编辑变得更加便捷和高效。然而，技术的双刃剑特性也带来了数字内容取证领域新的挑战。生成式大模型能够生成逼真的文本、图像、音频和视频，很可能被广泛用于制造虚假信息、深度伪造等恶意目的，对社会秩序和信息安全构成威胁。在大模型背景下，取证工作变得更加复杂，需要更高级别的技术手段来应对伪造技术的不断进步。为了应对网络虚假信息和大模型生成的高质量伪造内容，本报告介绍多项关键技术和一套整体的数字内容取证技术方案。本报告聚焦于传统图像篡改的检测和取证、人像深度伪造的检测，以及最新的 AIGC 图像和视频的检测，并针对已在网络上广泛传播虚假信息进行检测和事实性验证。面向大模型生成内容，还前瞻性地从源头入手，针对大模型进行知识编辑和输出内容的限制。这些研究从泛化性、可解释性、生成对抗博弈等多角度进行探索，取得了显著的成果，为在大模型背景下保障数字内容的真实性与可信性提供了重要方法和思路。

讲者简介：谭铁牛，研究员，博士生导师。现为南京大学党委书记、全国政协常委、中国科学院自动化研究所模式识别实验室主任、中国图象图形学学会名誉理事长，是中国科学院院士、英国皇家工程院外籍院士、发展中国家科学院院士、巴西科学院通讯院士、国际模式识别学会会士、国际电子电气工程师学会会士、中国图象图形学学会会士、中国人工智能学会会士。谭铁牛博士主要从事图像处理、计算机视觉和模式识别等相关领域的研究工作。主持过一批由国家基金委、国家 973 计划、863 计划、国际合作计划、国家重点研发计划等资助的科研项目。现已出版编著和专著 18 部，并在主要的国内外学术期刊和国际学术会议上发表论文 700 多篇，获准和申请发明专利 100 多项。曾获中国青年五四奖章、中国青年科技奖、国家技术发明二等奖、国家自然科学基金二等奖和国家科技进步二等奖和国际模式识别领域最高奖——傅京孙奖。曾任中央人民政府驻香港特别行政区联络办公室副主任、中国科学院副院长、中国人工智能学会副理事长、中国自动化学会副理事长、中国计算机学会副理事长、中国图象图形学学会理事长、国际模式识别学会第一副主席、IEEE 生物识别理事会主席。

报告主持人



简介：周杰，教授，博士生导师，IAPR Fellow，国家杰出青年科学基金获得者，国家基金委创新研究群体学术带头人，清华大学党委委员，全国重点实验室主任，兼任教育部高等学校自动化类专业教学指导委员会主任委员、国务院学位委员会学科评议组成员兼控制科学与工程评议组秘书长、人工智能领域国际顶级期刊《IEEE T-PAMI》编委（Associate Editor）、《模式识别与人工智能》副主编等。曾任清华大学副教务长、研究生院院长、国家卓越工程师学院执行院长、校学位评定委员会副主席、自动化系主任等职。长期从事模式识别与计算机视觉的学术研究，发表论文 300 多篇，其中 IEEE 期刊论文 130 多篇（含顶级期刊 IEEE T-PAMI 长文 39 篇）。以第一完成人获国家技术发明二等奖和中国专利银奖。指导的博士生获得全国优秀博士论文提名、中国人工智能学会优秀博士论文 5 人次。

大会报告（二）



报告人：佐藤洋一

报告题目：Understanding Egocentric Visual Attention and Actions

报告摘要： For a comprehensive understanding of human behavior, it is essential to know both their actions and the focus of their attention during various activities. In this talk, I will present our efforts in studying human visual attention and actions from first-person videos, shedding light on how individuals interact with their environment from an egocentric perspective.

专家简介： Yoichi Sato is a professor at the Institute of Industrial Science at the University of Tokyo. He received his B.S. degree from the University of Tokyo in 1990, and his MS and PhD degrees in robotics from the School of Computer Science at Carnegie Mellon University in 1993 and 1997, respectively. His research covers various topics in computer vision from object appearance analysis to human behavior understanding, such as egocentric vision, gaze sensing, physics-based vision, and reflectance analysis. He has been involved in journal editorial work, including IEEE Transactions on Pattern Analysis and Machine Intelligence, International Journal of Computer Vision, and Computer Vision and Image Understanding. He has also held key roles in organizing major conferences such as CVPR 2023 General Co-Chair, ICCV 2021 Program Co-Chair, ECCV 2012 Program Co-Chair, ACCV 2018 General Co-Chair, and ACCV 2016 Program Co-Chair.

报告主持人



简介： 查红彬，北京大学智能学院教授，机器感知与智能教育部重点实验室主任。主要从事计算机视觉与智能人机交互的研究，在三维视觉几何计算、三维重建与环境几何建模、传感器即时定位与地图构建等方面取得了一系列成果。出版学术期刊及国际会议论文 350 多篇，其中包括 IEEE T- PAMI, IJCV, IEEE T-VCG, IEEE T-RA 等国际期刊以及 ICCV, ECCV, CVPR 等国际学术会议论文 150 多篇。

大会报告（三）



报告人：王耀南(湖南大学)

报告题目：大模型驱动的具身智能人形机器人与展望

报告摘要：人形机器人是模仿人的形态运动和功能、与人交流、模仿人类外观和行为的高级通用智能机器人，可以辅助或者替代人类执行危险、肮脏、重复环境下多种类型任务。本文分析了大模型驱动的人形机器人技术进展，介绍了人形机器人研究背景与意义、国内外研究现状、大模型驱动的人形机器人关键技术，如大型自然语言模型、多模态视觉语言模型、具身智能多模态大模型等，以及发展趋势与展望，并举例介绍了具身智能人形机器人是国际公认的机器人技术集大成者和科技竞争的制高点，可推广应用于工业制造、国防安全、智能服务和智慧医养等行业，具有广阔的前景和巨大的潜力。提出大模型技术的发展为人形机器人感知识别、认知决策、人机交互、自主学习、多机协同规划、行为操作控制系统注入语言理解、视觉泛化、常识推理等关键能力和实现方法，有望推动具身智能人形机器人技术与应用的新一轮发展。

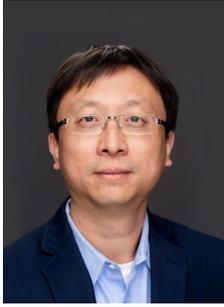
专家简介：王耀南，中国工程院院士，机器人技术与智能控制专家，湖南大学教授，机器人视觉感知与控制技术国家工程研究中心主任；中国科协委员、中国图象图形学学会理事长、中国自动化学会会士、中国计算机学会会士、中国人工智能学会会士；国家863智能机器人领域专家；主持完成国家科技计划重大专项、重大科研仪器研制项目、国家863重大项目、国家自然科学基金重大项目、欧盟第五框架国际合作项目、德国杰出洪堡学者项目等；技术成果获国家技术发明二等奖1项、国家科技进步二等奖5项、中国发明创业特等奖1项、省部级科技奖励一等奖11项、国际IEEE机器人与自动化领域“工业应用最高奖”1项；发表高水平论文200余篇，出版著作15部，发明专利90余项。

报告主持人



个人简介：林宙辰，北京大学教授，IAPR/IEEE/CSIG Fellow，国家杰青，中国图象图形学学会机器视觉专委会主任，中国自动化学会模式识别与机器智能专委会副主任。研究领域为机器学习、数值优化。发表论文230余篇，英文专著2本。获2020年度CCF科学技术奖自然科学一等奖。多次担任机器学习与人工智能顶级会议CVPR、ICCV、NIPS/NeurIPS、ICML、IJCAI、AAAI和ICLR的领域主席，曾任IEEE T. Pattern Analysis and Machine Intelligence编委，现任ICPR 2022程序共同主席、ICML 2022资深领域主席，International J. Computer Vision、Optimization Methods and Software编委。

大会报告（四）



报告人：马毅(香港大学)

报告题目：透明且一致的深度表征学习：从黑盒到白盒，从开环到闭环

报告摘要：在本次报告中，我们从压缩数据的编码和解码的角度，对过去十年深度神经网络的实践进行了系统的解释。我们认为，学习（或智能）的最根本目标是学习到一个紧凑且有结构的数据分布表示。最终学习到的表示的优劣可以通过一个信息增益的原则性量度来评估，该量度可以通过学习特征的（有损）编码率计算得出。我们主张，对这一目标的展开迭代优化提供一个统一的白盒解释，可以解释几乎所有过去和当前广泛应用于人工智能实践的深度神经网络，包括 ResNets 和 Transformers。我们将通过令人信服的理论 and 实证证据表明，数学上可解释、实践上有良好表现且语义上有意义的深度网络现在已触手可及。此外，我们的研究表明，为了使学习到的表征正确且一致，必须闭合编码和解码网络的循环，而不是像当前的做法那样，将它们作为独立的开环网络进行端到端的训练。最重要的是，这一新框架揭示了开发下一代高效的自主学习架构和系统的广阔而光明的未来，这些系统可以真正模拟记忆创建的计算机制。

专家简介： Yi Ma is a Chair Professor in Artificial Intelligence, the inaugural director of the Data Science Institute, and the head of the Computer Science Department of the University of Hong Kong since 2023. His research interests include computer vision, high-dimensional data analysis, and integrated intelligent systems. Yi received his two bachelor's degrees in Automation and Applied Mathematics from Tsinghua University in 1995, two master's degrees in EECS and Mathematics in 1997, and a PhD degree in EECS from UC Berkeley in 2000. He has been on the faculty of UIUC ECE from 2000 to 2011, the principal researcher and manager of the Visual Computing group of Microsoft Research Asia from 2009 to 2014, and the Executive Dean of the School of Information Science and Technology of ShanghaiTech University from 2014 to 2017. He joined the faculty of UC Berkeley EECS in 2018. He has published over 65 journal papers, 145 conference papers, and 3 textbooks on computer vision, generalized PCA, and high-dimensional data analysis. He received the NSF Career award in 2004 and the ONR Young Investigator award in 2005. He also received the David Marr prize in computer vision from ICCV 1999 and best paper awards from ECCV 2004 and ACCV 2009. He has served as the Program Chair for ICCV 2013 and the General Chair for ICCV 2015. He is a Fellow of IEEE, ACM, and SIAM.

报告主持人



简介：库尔班·吾布力，博士、教授、博士生导师。中国计算机学会（CCF）和中国图象图形学学会（CSIG）理事、CCF 乌鲁木齐的副主席、NOI 新疆竞赛委员会主席、GESP 和 PTA 新疆大学考点和服务中心的负责人、自治区教育工委优秀共产党员、新疆维吾尔自治区天山英才科技创新领军人才、自治区青少年科普专家。中国自动化学会模式识别与机器智能专业委员会(CAA-PRMI) 副秘书长、CCF、CAA、CSIG、CAAI 等学会在人工智能领域的多个专业委员会委员、CCF/CSIG 杰出会员、IEEE, IAPR, IAENG 会员、新疆大学计算机学科群“双一流”科，研平台一大规模场景实时三维重建与理解平台负责人。主要研究方向为图像处理、模式识别、计算机视觉、自然语言处理等。主持国家重点研发课题 1 项、国家自然科学基金 4 项，省部级和其他项目 10 余项，在国内外期刊和会议上发表论文 300 余篇，出版学术专著 3 部，专利 10 余项，计算机软件著作权 60 余项，获得中国图象图形学科技进步二等奖、自治区自然科学优秀论文二等奖等省部级以上奖项 20 余项。

特邀报告信息及讲者简介

特邀报告（一）



报告人：王立君

报告题目：开放条件下的视觉识别

报告摘要：随着多模态感知与大模型技术的飞速发展，视觉识别也从可控封闭环境迈向更加开放的场景，一方面体现为输入数据模态具有不确定，现有模型在测试时难以有效应对任意输入模态带来的挑战；另一方面，目标类别具有开放性，测试阶段待识别的类别与训练集类别不一致、或需要额外的逻辑推理才可以确定，极大程度提升了视觉识别的难度。针对上述问题，本报告将从输入模态、开放类别字典和因果推理三个层面，介绍团队在开放视觉识别领域最新的研究成果，探讨如何强化视觉识别模型对模态的鲁棒性和类别的泛化性，以及如何借鉴多模态大模型技术面向视觉识别任务提升开放环境下的推理能力。

讲者简介：王立君，大连理工大学未来技术/人工智能学院特聘副教授，博士生导师，国家自然科学基金获得者，主要研究方向聚焦于图像深度估计、目标识别、检测与跟踪等。主持国家自然科学基金重点、面上和青年项目，入选人社部“博士后创新人才支持计划”和大连市“科技人才创新支持计划”，在本领域顶级学术会议和期刊发表论文 40 余篇，谷歌学术总引用 7900 余次。相关研究成果获得辽宁省科技进步一等奖，中国图象图形学会自然科学二等奖，教育部自然科学二等奖，中国图象图形学学会优秀博士论文奖，以及辽宁省优秀博士论文奖。连续三年获得 VOT 国际视觉跟踪竞赛冠军。担任多个国际会议和期刊审稿人，Valse 执行委员，CCF-CV 与 CSIG-MV 专委会执行委员等。

特邀报告（二）



报告人：彭宇新

报告题目：细粒度多模态运动分析

报告摘要：运动是生命的基本要素，是现实世界的客观描述。与粗粒度、单模态的运动分析不同，细粒度、多模态的运动分析描述真实世界更准确，同时更具挑战性。细粒度多模态运动分析旨在利用文本、视频、3D 姿态等不同模态信息的推理能力对目标动作进行细粒度识别、定位、评价、重建等，关键科学问题是如何精准辨识时空边界模糊的细粒度子动作，如何利用文本信息对人体动作进行精细重建等。可广泛应用于智能安防、智慧医疗、智能体育、智慧传媒等领域。本报告将介绍我们最新的 6 篇 CVPR 2024 论文，涵盖了我们在细粒度多模态运动分析的近期工作进展，包括细粒度动作质量评价、文本驱动细粒度时空动作定位、文本驱动的三维人体姿态估计等，并对未来发展方向与趋势进行讨论与展望。

讲者简介：彭宇新，北京大学二级教授、博雅特聘教授、国家杰出青年科学基金获得者、国家万人计划科技创新领军人才、科技部中青年科技创新领军人才、863 项目首席专家、中国人工智能产业创新联盟专家委员会主任、中国工程院“人工智能 2.0”规划专家委员会专家、中国电子学会会士、中国人工智能学会会士、中国图象图形学学会会士、副秘书长、提名与奖励委员会副主任、北京图象图形学学会副理事长。主要研究方向为跨媒体分析、计算机视觉、机器学习、人工智能。以第一完成人获 2016 年北京市科学技术奖一等奖和 2020 年中国电子学会科技进步奖一等奖，2008 年获北京大学宝钢奖教金优秀奖，2017 年获北京大学教学优秀奖。主持了 863、国家自然科学基金重点等 30 多个项目，发表论文 200 多篇，包括 ACM/IEEE Trans 和 CCF A 类论文 100 多篇。多次参加由美国国家标准技术局 NIST 举办的国际评测 TRECVID 视频样例搜索比赛，均获第一名。主持研发的跨媒体互联网内容分析与识别系统已经应用于公安部、工信部、国家广播电视总局等单位。担任 IEEE TMM、TCSVT 等期刊编委。

特邀报告（三）



报告人：王兴刚

报告题目：文本提示“检测分割一切”模型研究

报告摘要：Meta 公司开发的“分割一切”模型（SAM）在全世界获得了广泛的关注和应用，但在 SAM 和 SAM 2 中都只支持点、包围盒等形式的交互提示，无法做到自动化的文本提示的分割一切物体。针对这个难题，我们从视觉特征和文本特征的高效交互出发，在视觉基础模型和文本基础模型的基础上，设计了大规模弱监督预训练、早期多模态特征融合等有效策略，分别开发出了 YOLO-World 模型（CVPR 24, 4.1k github stars）和 EVF-SAM 模型（arXiv:2406.20076, 225 github stars），取得了业内领先的检测、分割一切目标的性能。

讲者简介：王兴刚，华中科技大学电信学院教授博导，国家级青年人才，现任 Image and Vision Computing 期刊（Elsevier, IF 4.2）共同主编。主要从事基础模型、视觉表征学习、目标检测分割跟踪等领域研究、在 IEEE TPAMI、IJCV、CVPR、ICCV、NeurIPS 等顶级期刊会议发表学术论文 60 余篇，谷歌学术引用 30000 余次，其中一作/通讯 1000+ 引用论文 5 篇，入选 Elsevier 2023 中国高被引学者。担任 CVPR, ICCV, ICIG 等会议领域主席，Machine Vision and Application 等期刊编委。入选了中国科协青年人才托举工程，获湖北青年五四奖章、CSIG 青年科学家奖，吴文俊人工智能优秀青年奖，CVMJ 2021 最佳论文奖，湖北省自然科学二等奖等，指导学生获 2022 年全国“互联网+”大赛金奖、2023 年挑战杯“揭榜挂帅”专项赛全国一等奖。

特邀报告（四）



报告人：张兆翔

报告题目：视觉物体检测的新趋势和新范式

报告摘要：视觉物体检测是计算机视觉与模式识别领域的重要问题，在过去十余年间伴随着深度学习的兴起取得重要进展。伴随着多模态大模型、生成式模型、世界模型等新兴技术的兴起和自动驾驶、机器人等关键应用的驱动，视觉物体检测的方向发展呈现出新趋势，视觉物体检测的方法创新呈现出新范式。本报告在回顾视觉物体检测的发展基础上，着重介绍本团队在视觉物体检测领域做的一系列代表性工作，并以自动驾驶等应用场景，生成式世界模型等方法驱动为例，探讨视觉物体检测的创新发展。

讲者简介：张兆翔，博士，研究员，博士生导师，长江学者，中国科学院自动化研究所模式识别实验室常务副主任，中国科学院大学岗位教授，曾获得北京市科学技术奖科技进步奖一等奖（排名第一）。张兆翔博士的研究兴趣包括：模式识别、具身智能、智能体学习，在 IEEE T-PAMI、IJCV、JMLR、National Science Review 等顶级期刊与 CVPR、ICCV、ECCV、NIPS、AAAI、IJCAI 等顶级会议发表论文 200 余篇，授权专利 30 余项，承担了国家自然科学基金重点项目、重点国际(地区)合作研究项目、企业联合重点支持项目、国家重点研发项目等多项国家级科研项目，多次担任 CVPR、ICCV、NeurIPS 等顶会 Area Chair。

特邀报告（五）



报告人：王大轶

报告题目：基于能力定量表征的序列图像自主导航技术

报告摘要：深空探测任务对空间飞行器全自主运行技术提出了迫切需求，自主导航与自主诊断重构是其中两大关键核心，也是实现全自主运行的前提和保障。报告人针对资源强受限的空间飞行器，以观测能力的定量表征为理论创新突破口，以从序列图像中挖掘导航状态信息为关键技术突破口，提出了基于观测能力定量表征的序列图像自主导航技术，为实现空间飞行器安全可靠自主运行做出重要贡献。

讲者简介：王大轶（北京空间飞行器总体设计部科技委主任、研究员）是国家杰青、国防卓青、万人领军，973 项目（自主诊断重构）和 173 重点项目（自主导航）技术首席专家，长期从事观测、诊断和重构能力定量表征理论方法以及空间飞行器全自主运行技术研究，为我国探月工程和首次火星探测等任务的圆满完成做出重要贡献。获国家技术发明二等奖、国家科技进步特等奖，以及全国创新争先奖、何梁何利基金科技创新奖、首届国家工程师奖等，被授予“国家卓越工程师”称号。

特邀报告（六）



报告人：刘焯斌

报告题目：高真实感 3D 高斯数字化身

报告摘要：近期，3D 高斯泼溅（Guassian Splatting）作为一种新的可微渲染技术，逐渐取代其他可微渲染方法成为学术界和产业界的研究热点。借助 3D 高斯泼溅在 3D 场景表达和渲染的优势，介绍报告人在 3D 数字重建与生成方面的最新研究工作，包括高保真的头部与身体数字化身生成、简易-快速高效的人头 3D 高斯化身、融合视频生成机制的 3D 高斯数字化身、基于动作生成的高斯数字化身交互等相关研究工作。

讲者简介：刘焯斌，清华大学长聘教授，国家杰青基金获得者。研究方向为三维视觉、数字人重建、3D 生成与交互。发表计算机视觉/计算机图形学领域顶会顶刊近百篇。多次担任 CVPR、ICCV、ECCV 领域主席，担任 IEEE TVCG、CGF 编委，中国图象图形学会三维视觉专委会副主任。获 2012 年国家技术发明一等奖（排名 3），2019 年中国电子学会技术发明一等奖（排名 1）。

特邀报告（七）



报告人：常建龙

报告题目：通用数据预训练与高效微调方法研究

报告摘要：人工智能的发展和应用一直面临着智能孤岛的问题，即各种数据中的知识只能通过不同的模型学习，而不能累积学习。不同于人工智能，人类智能没有智能孤岛现象，可以从海量任务和数据中累积学习，这就是为什么人类能够根据他们过去的经验完成他们从未遇到过的任务。该报告聚焦于该问题，从数据、架构和优化出发实现预训练模型的累积式知识学习，并从通过统一各类微调方案解提升预训练模型的应用效率。

讲者简介：华为 AI 技术专家、百度奖学金、中国人工智能协会吴文俊优秀博士学位论文奖、中科院优秀博士学位论文奖、中科院院长奖获得者。主要研究通用 AI 算法和商业落地，五十余项研究成果发表于 CSUR、IEEE T-PAMI、IEEE T-IP、NeurIPS、CVPR 等国际顶级 AI 期刊会议，并成功应用于华为云为千行百业赋能。

特邀报告（八）



报告人：杜博

报告题目：多模态大模型发展及垂域赋能

报告摘要：多模态大模型作为人工智能领域的重要进展，通过整合学习文本、视觉等多模态数据信息，极大地提升了模型的多模态理解和人机交互能力。随着大模型技术的不断发展，多模态大模型也有望助力赋能各行业发展。本报告将探讨多模态大模型的发展历程、应用场景与研究挑战，以及介绍团队从通用基础大模型到医疗、遥感垂直领域多模态大模型的研究成果，分析多模态大模型在特定垂直领域的应用潜力和赋能效果，并为未来的大模型技术发展和赋能应用提供前瞻性视角。

讲者简介：杜博，武汉大学弘毅特聘教授，国家自然科学基金杰出青年科学基金、优秀青年科学基金获得者，湖北省杰青。现任国家多媒体软件工程技术研究中心主任，武汉大学发展规划与学科建设办公室主任、计算机学院院长，武汉大学人工智能研究院常务副院长，多媒体网络通信工程湖北省重点实验室主任。主要从事计算机视觉和人工智能等方面的研究工作。近五年主持和参与国家重点研发计划项目等纵向科研项目 30 余项，发表 ESI 高被引或热点论文 23 篇，出版著作 3 部，授权国家发明专利 35 项。获得 2024 年湖北省首届青年创新奖（序 1），2019 年湖北省自然科学一等奖（序 1）；2020、2021 年中国国际高新技术成果交易会优秀产品奖；2020 年湖北省首届专利银奖；2019-2023 年科睿维安全全球高引学者；2020-2023 年 Elsevier 中国高被引学者；多次获多项人工智能领域顶级会议最佳论文奖以及人工智能和图像处理类全球竞赛冠军。

专题论坛日程

专题论坛：视觉基础模型及其前沿应用

(论坛主席：罗斌，安徽大学计算机科学与技术学院教授，
范登平，南开大学教授)

(10月18日 10:00-12:00, 地点:报告厅)

时间	报告嘉宾姓名、单位、及联系信息	报告题目
10月18日		
10:00-10:24	郑伟诗 中山大学 wszheng@ieee.org	具身视觉感知与学习
10:24-10:48	王兴刚 华中科技大学 xgwang@hust.edu.cn	低复杂度视觉基础模型研究
10:48-11:12	叶茫 武汉大学 yemang@whu.edu.cn	多模态基础模型驱动的图文跨模态检索
11:12-11:36	任文琦 中山大学 renwq3@mail.sysu.edu.cn	低质量与对抗环境下的目标检测
11:36-12:00	王文海 上海人工智能实验室青年科学家 wangwenhai@pjlab.org.cn	视觉语言大模型的研究与应用：通过开源套件缩小与商用多模态大模型的性能差距

论坛介绍

在大规模基础模型取得重大进展的推动下，计算机视觉领域经历了深刻的变革。计算机视觉领域的变革浪潮前沿是视觉模型，例如 Vision Transformer (ViT)，通过自注意力机制来理解图像，引领了图像理解范式的转变。同时，跨语言图像预训练 (CLIP) 框架提出了一种新的方法，通过利用图像-文本对数据集来促进视觉-语言理解。此外，Segment Anything Model (SAM) 和 GPT-4V 为视觉模型赋予了交互式语义分割的功能。这些进展不仅突显基础模型理论的重要性，而且还提供了具体的实现路径，重塑计算机视觉的格局。然而，视觉基础模型的研究方兴未艾，如何在实际应用场景中产生更大的价值，值得我们进一步探索。基于此，中国图象图形学学会主办的英文学术期刊 Visual Intelligence 将在本次大会上组织“视觉基础模型及其前沿应用”专题论坛，旨在将领域内相关研究者聚集在一起，交流目前视觉基础模型研究和技术的最新进展和未来发展趋势，对其中一些共性、难点问题进行交流 and 深入探讨。

论坛主席



个人简介：罗斌(luobin@ahu.edu.cn)，安徽大学计算机科学与技术学院教授、博士生导师，安徽省学术技术带头人，安徽省模范教师。现任中国图象图形学学会常务理事、安徽省计算机学会监事长。曾受聘英国电信公司短期研究员、美国佛罗里达技术学院客座教授、澳大利亚新南威尔士大学客座研究员、新加坡南洋理工大学客座教授、英国 Stirling 大学客座教授、英国 York 大学博士后研究员等。研究方向为模式识别与数字图像处理。代表性成果发表于 IEEE TPAMI 和 CVPR 等国际期刊和会议。任 Visual Intelligence 副主编和《安徽大学学报（自然科学版）》主编。



个人简介：范登平(fdp@nankai.edu.cn)，南开大学教授、博士生导师，计算机系主任、国家级四青人才。代表性论文：1 篇 NeurIPS workshop 最佳论文、2 篇 CVPR 最佳论文提名、5 篇热点论文、10 篇 TPAMI，谷歌引用 2.1 万，H 指数 53。担任 IEEE TIP、Visual Intelligence 等国际期刊编委，CVPR、NeurIPS 等国际顶会领域主席，入选斯坦福全球前 2% 顶尖科学家榜单。

特邀讲者简介



特邀讲者 1: 郑伟诗 中山大学教授

报告题目: 具身视觉感知与学习

报告摘要: 面向物理实体机器人的视觉行为感知与学习是具身智能研究的重要一环。本次报告将汇报我们研究组在通用机器人自由抓取建模上所展开的相关研究,包括机器人6DOF 抓取建模、机械灵巧手抓取建模、动态抓取与多机器人协作、机器人行为质量评估等初步研究成果。

个人简介: 郑伟诗(wzheng@ieee.org), 中山大学计算机学院教授、副院长, 教育部“长江学者奖励计划”特聘教授, 英国皇家学会牛顿高级学者, 教育部机器智能与先进计算重点实验室主任。他长期研究协同与交互分析理论与方法, 解决人体建模和机器人行为的视觉计算问题。担任 IEEE TPAMI、Visual Intelligence 等期刊的编委。主持承担国家级重点类项目和人才项目 5 项、广东省自然科学基金委卓越青年团队(负责人)项目等。获国家教学和省部级科技奖励 6 项。



特邀讲者 2: 王兴刚 华中科技大学教授

报告题目: 低复杂度视觉基础模型研究

报告摘要: Transformer 在视觉和多模态基础模型领域已取得巨大成功,但其计算复杂度随输入序列长度的增加而呈二次增长。因此,探索低复杂度的视觉和多模态基础模型成为一个核心问题。本报告将介绍我们利用 Mamba、Gated Linear Attention 等新机制构建低复杂度视觉基础模型的初步研究,以及如何运用这些新型网络实现计算高效且表征能力强的视觉理解和内容生成模型。

个人简介: 王兴刚(xgwang@hust.edu.cn), 华中科技大学电信学院教授、博士生导师,入选国家级青年人才计划,现任 Image and Vision Computing 期刊(Elsevier, IF 4.2)共同主编。主要从事基础模型、视觉表征学习、目标检测分割跟踪等领域研究、在 IEEE TPAMI、IJCV、CVPR、ICCV、NeurIPS 等顶级期刊会议发表学术论文 60 余篇,谷歌学术引用 3 万余次,其中一作/通讯 1000+引用论文 5 篇,入选 Elsevier 2023 中国高被引学者。担任 CVPR、ICCV、ICIG 等会议领域主席,Machine Vision and Application 等期刊编委。入选中国科协青年人才托举工程,获湖北青年五四奖章、CSIG 青年科学家奖,吴文俊人工智能优秀青年奖,CVMJ 2021 最佳论文奖,湖北省自然科学二等奖等,指导学生获 2022 年全国“互联网+”大赛金奖、2023 年挑战杯“揭榜挂帅”专项赛全国一等奖。



特邀讲者 3: 叶茫 武汉大学教授

报告题目: 多模态基础模型驱动的图文跨模态检索

报告摘要: 图文跨模态检索任务是连接视觉内容与文本信息的重要桥梁, 如何将通用多模态基础模型迁移适配到下游任务具有重要意义。本次报告将分享我们团队基于通用 CLIP 模型在下游的图文跨模态检索上的研究进展, 主要包括以人为中心的图文检索和多模态检索, 通用图文检索和视频文本检索等方向, 旨在实现更精准高效的信息检索。

个人简介: 叶茫(yemang@whu.edu.cn), 武汉大学计算机学院教授、博士生导师, 国家级高层次青年人才, 中国科协青年托举人才, 湖北省高层次人才。曾任阿联酋起源人工智能研究院研究科学家, 美国哥伦比亚大学访问学者。主要研究方向多模态检索、多模态语义理解、联邦学习等, 以第一/通讯作者发表 CCF-A 类论文 60 余篇, 谷歌学术引用 9000 余次, 1 篇论文入选湖北省优秀科技论文。受邀担任 CCF-A 类期刊 IEEE TIFS (中科院一区) 等国际 SCI 期刊编委, 多次受邀担任顶级会议 CVPR、ACM MM、NeurIPS、ICLR、ECCV 领域主席等学术职务。



特邀讲者 4: 任文琦 中山大学教授

报告题目: 低质量与对抗环境下的目标检测

报告摘要: 本报告聚焦低质量视觉内容增强与感知对抗主题。针对图像视频在成像过程中常见的模糊、雨雾、低分辨率及低动态范围等质量退化问题,深入研究了多种视觉基础模型在视觉内容清晰化上的可行性,验证了新颖 Transformer 和状态空间模型在底层视觉恢复任务的有效性。提出一种新的图像恢复算法,在提升图像的清晰度的同时充分融合了下游目标检测效果的增益考量。通过对干净标签进行对抗攻击能够生成既与原始样本相近又有助于目标检测的伪标签,从而确保在增强图像质量的同时提升目标检测的准确度。此外,针对视觉对抗攻击这一日益严重的威胁,进一步提出一种无需先验知识或额外训练的对抗补丁定位与移除方法。该方法能够应对各种对抗攻击模式,兼容不同预训练的目标检测器,显著提升了防御能力。提出的方法在多种退化场景和攻击类型下均能显著提升目标检测的准确性。

个人简介: 任文琦(renwq3@mail.sysu.edu.cn),中山大学“百人计划”引进人才,教授,主持国家自然科学基金优秀青年基金,广东省重大人才工程青年拔尖人才。长期从事多媒体内容安全、目标增强与感知、计算机视觉等领域的教学科研工作。在人工智能和计算机视觉领域发表 80 余篇中科院一区 and CCF-A 类期刊和国际顶会议论文。截至目前,Google Scholar 引用 14000 余次,包括 ESI 高被引论文 7 篇,第一作者论文单篇最高被引 1800 余次,入选 2022/2023 年爱思唯尔中国高被引学者。担任 CVPR、NeurIPS、ICLR 等计算机视觉与人工智能国际会议的领域主席,《中国图像图形学报》等期刊编委。主持国家自然科学基金优青、面上、国际合作与交流项目,深圳市优青、面上项目,华为、腾讯公司资助项目等 20 余项。入选 2018 微软亚洲研究院“铸星计划”、2022 百度全球高潜力 AI 华人青年学者。曾获 2019 年中国计算机学会优博奖、2022 年吴文俊人工智能优秀青年奖、2023 年中国图象图形学学会自然科学一等奖(第三完成人)。



特邀讲者 5: 王文海 上海人工智能实验室青年科学家

报告题目: 视觉语言大模型的研究与应用: 通过开源套件缩小与商用多模态大模型的性能差距

报告摘要: 随着大语言模型的兴起, 多模态大模型也取得了显著进步, 推动了复杂的视觉语言对话和交互, 弥合了文本与视觉信息之间的鸿沟。然而, 现有的开源模型与商用闭源模型 (如 GPT-4o 和 Gemini 1.5 Pro) 相比, 在模型性能上仍存在显著差距。本报告将探讨图文多模态大模型的基本原理和技术, 探索如何利用开源套件构建强大的多模态大模型, 研究如何通过渐进式策略扩展开源多模态模型的性能边界, 以缩小开源模型与商业闭源模型在多模态理解方面的能力差距。

个人简介: 王文海(wangwenhai@pjlab.org.cn), 上海人工智能实验室青年科学家。南京大学博士, 香港中文大学博士后, 上海人工智能实验室“书生”系列视觉和多模态模型核心开发者。研究方向为视觉基础模型, 在 IEEE TPAMI、CVPR、ICCV、ECCV、ICLR、NeurIPS 等国际会议和期刊发表论文 40 余篇, 其中 20 篇为一作/共同一作/通信作者。研究成果获得了超过 2 万次引用, 单篇最高引用超 3000 次。研究成果分别入选 CVPR 2023 最佳论文, 世界人工智能大会青年优秀论文奖, CVMJ 2022 最佳论文提名奖, 两次入选 ESI 高被引论文 (前 1%) 和热点论文 (前 0.1%), 6 次入选 Paper Digest CVPR、ICCV、NeurIPS、ECCV 年度十大最具影响力论文。入选斯坦福大学 2023 年度全球前 2% 顶尖科学家, 中国图象图形学学会优博提名, 世界人工智能大会云帆奖。担任 Visual Intelligence 编委, IJCAI 2021 的高级程序委员会委员, 以及多个顶级国际期刊和会议的审稿人。

本地主持人



简介：钱育蓉，博士，教授，博士生导师，现任新疆大学计算机科学与技术学院党委书记、院长，教育部“丝路多语言认知计算”一带一路国际联合实验室常务副主任，新疆科协委员、新疆欧美同学会（新疆留学人员联谊会）理事、中国计算机学会高级会员，入选自治区杰出青年、“天山英才”和“万人计划”后备人选，系国家级一流专业“软件工程”负责人、国家级一流课程“数据库原理与技术”负责人、自治区级创新团队“面向农业的天地协同水资源时空精准调度研究及应用创新团队”负责人。曾获国家教学成果二等奖、新疆维吾尔自治区科技进步一等奖、新疆维吾尔自治区教学成果奖一等奖和新疆维吾尔自治区自然科学优秀论文奖等奖项，曾获全国宝钢优秀教师奖、自治区优秀共产党员，全国“互联网+”金奖优秀指导教师、国家大学生创新计划项目指导教师。长期从事人工智能（大数据、图像处理）、网络计算和农业遥感领域研究，主持国家自然科学基金、国防科工委高分遥感应用重大专项和省部级各类科研项目 20 余项，近五年以第一作者/通讯作者在国内知名核心刊物发表论文近 100 篇，其中 SCI/EI 收录论文 42 篇，中文核心期刊论文 57 篇，授权发明专利 3 项，软件著作权三十余项。



简介：艾斯卡尔·艾木都拉，男，1972 年 2 月生，博士（后），教授，博士生导师，教育部“新世纪优秀人才支持计划”入选者，首届《新疆网络安全人才奖》获得者，新疆第七届《十大杰出青年》，新疆维吾尔自治区“教书育人”先进个人，全国“宝钢教育基金”优秀教师奖获得者，自治区重点（高峰）学科“计算机科学与技术”方向带头人，全国青联委员，中国中文信息学会语音学分会副主任委员，中国语言学会语音学分会委员，自治区青联青年科技工作者协会副会长，现任新疆大学未来技术学院副院长、新疆多语种信息技术重点实验室副主任等职务。主要从事语音识别与合成、图像文字识别、智能信息检索、信息内容安全、自然语言处理、机器翻译、智能无人系统等方面的研究工作，已发表学术论文 250 余篇，发明专利 20 余项，以第一完成人身份获得自治区科技进步奖二等奖 2 项。



简介：吕小毅，教授、博士生导师、2010 年博士毕业于西安交通大学，新疆自治区高层次引进人才，现任新疆大学科技处副处长（主持工作），新疆信号检测与处理重点实验室副主任，新疆抗癌协会肿瘤人工智能专业委员会主任委员，中国计算机学会生物信息学专委会委员，中国人工智能学会生物信息与人工生命专委会常委，中国抗癌协会肿瘤人工智能专业委员会委员，中国计算机学会乌鲁木齐分部监督委员会主席。自治区天山雪松科技创新领军人才，自治区杰出青年基金获得者。主持科技部重点研发课题项目，国家自然科学基金项目、自治区揭榜挂帅项

目、自治区重点研发等项目 20 余项，以第一作者或通信作者身份发表 SCI 论文 180 余篇，获授权国家发明专利 20 余项。



简介：刘锦，男，中南大学计算机学院副教授，新疆大学软件学院副院长，博/硕士研究生导师，湖南省普通高校青年骨干教师，湖南省优秀研究生导师团队骨干，CCF 高级会员，CSIG 脑图谱专业委员会委员，CCF 生物信息学专委会委员，湖南省数字医学学会医学智能影像分会常务委员，湖南省神经科学学会帕金森病及运动障碍疾病专委会委员，主要研究方向是人工智能、医疗及健康大数据分析、健康监测智能设备研发等。目前已在 IEEE TMI、MedIA、IEEE/ACM TCBB、ACL、AAAI、ICML 等国际学术期刊或会议上发表高水平学术论文 70 余篇，其中 1 篇 Featured Article，ESI Top 1% 高被引论文 6 篇，最佳论文奖 5 次，据 Google Scholar 统计显示总引用 3200 余次，单篇最高引用 500 余次，H 指数为 28，入选美国斯坦福大学发布的全球前 2% 顶尖科学家榜单（2023/2024），《Big Data Mining and Analytics》，《工程科学学报》等国内外期刊青年编委。主持或作为科研骨干参与国家重点研发计划、国家自然科学基金面上基金、湖南省重点领域研发计划等省部级科研项目 10 余项，荣获湖南省科学技术进步奖二等奖、湖南省医学会医学科学技术一等奖、中国研究型医院学会医学研究创新二等奖、湖南省优秀博士论文奖等多项学术奖励。



简介：阿力木江·艾沙，工学博士、教授，新疆大学党委教师工作部副部长。1997 年毕业于四川大学计算机系，分别于 2008 年和 2013 年获得新疆大学计算机应用硕士和博士学位。中国计算机学会（CCF）高级会员、中国图象图形学学会（CISG）会员、中国图形图象学学会文档分析与识别专业委员会（CISG-DIAR）委员。从事模式识别、自然语言处理、计算机视觉方面的研究。主持或参与国家自然科学基金项目 3 项，省部级和其他项目 6 项，在国内外期刊和会议上发表论文 70 余篇，出版学术专著 2 部、专利 2 项，计算机软件著作权 20 余项。



简介：夏威，工学博士，电子科技大学信息与通信工程学院副教授，新疆大学计算机科学与技术学院（网络空间安全学院）副院长（援疆）。分别于 2002 年、2005 年、2008 年获电子科技大学通信工程学士学位、信息与通信工程硕士学位、信号与信息处理博士学位。长期从事于自适应信号处理、阵列信号处理、雷达信号处理、人工智能等领域的研究，已在国内外学术期刊和国际学术会议发表论文 120 余篇，授权国家发明专利 20 余项，先后主持国家自然科学基金项目 4 项和多项省部级科研项目，并作为主研人员参加国家自然科学基金重点项目、国家重点研发计划等多项国家级、省部级以及横向科研项目。

专题论坛：女科学家论坛

(论坛主席：贲晔焯，山东大学教授，

陈俊颖，华南理工大学教授)

(10月18日 15:30-17:15, 地点：塔里木厅)

时间	报告嘉宾姓名、单位、及联系信息	报告题目
10月18日		
15:30-15:35	贲晔焯 山东大学 benxianye@126.com 陈俊颖 华南理工大学 jychense@scut.edu.cn	开场白
15:35-15:55	张艳宁 西北工业大学 ynzhang@nwpu.edu.cn jqyang@nwpu.edu.cn	以文为媒：高阶语义知识引导的视觉重建、感知与理解
15:55-16:15	董晶 中国科学院自动化研究所 jdong@nlpr.ia.ac.cn	数字内容生成与可信鉴伪研究浅析
16:15-16:35	杨欣 华中科技大学 xinyang2014@hust.edu.cn	视觉空间计算方法及应用
16:35-16:55	张铭津 西安电子科技大学 mjinzhang@xidian.edu.cn	跨域推理赋能的智能图像处理
16:55-17:15	王路 哈尔滨工程大学 wanglu2019@hrbeu.edu.cn	海杂波耦合情况下舰船尾迹检测技术

论坛介绍

在模式识别和计算机视觉领域，一代代优秀的女科学家们坚持不懈，凭着坚韧、细致、认真的工作态度，勇敢迎接科研领域的新挑战，发掘科研领域的新机遇，在科技攻关中彰显女性力量，突破障碍，取得显著成就。本论坛邀请了 5 位在模式识别和计算机视觉领域取得突出成绩的优秀女科学家们分享她们的创新研究成果，共同探讨女科学家的成长之路。

论坛主席



个人简介：贲晔(benxianye@126.com)，山东大学教授，博导，国家级青年人才，山东省杰青，山东省泰山学者青年专家，山东大学杰出中青年学者。研究方向为图像处理、模式识别。IEEE Senior Member，中国电子学会高级会员，中国电子学会信号处理分会委员，中国电子学会青年科学家俱乐部理事，山东省人工智能学会常务理事等。作为项目负责人，主持国家自然科学基金优秀青年科学基金项目、国家重点研发计划项目课题、山东省重大科技创新工程项目等 20 余项国家级和省部级项目。在 IEEE T-PAMI、IEEE T-IP、IEEE T-CSVT、IEEE T-MM、PR、CVPR 等国内外核心期刊会议上发表学术论文 100 余篇，授权国家发明专利 71 项，成果在宁夏银川监狱、内蒙古呼和浩特第三监狱和江西省赣州监狱的个人谈话、心理矫治与智能审讯系统以及青岛两个社区的视频监控系统上成功应用，服务了社会治理，为平安中国的建设做出重要贡献。获山东省自然科学奖二等奖（排名 1）、中国图象图形学学会石青云女科学家奖、山东省优秀研究生学位论文指导教师、青岛市巾帼建功标兵、青岛市巾帼科研之星、山东大学“三八红旗手”等称号。



个人简介：陈俊颖，华南理工大学教授、博士生导师，国家卓越工程师学院导师，大数据与智能机器人教育部重点实验室主任助理、智能医学图像处理与医疗机器人研究室负责人，香港大学博士，浙江大学学士（优秀毕业生），CCF 技术公益大使，CCF YOCSEF 广州学术秘书，CSIG 广州会员活动中心副秘书长，广州市人工智能学会理事。主要研究多源多模态特征融合、计算机视觉、智能机器人等，成果发表于 TPAMI、TNNLS、TMI、CVPR、SMC、IROS 等期刊会议，获得 2023 年广东省计算机学会青年科技奖、2023 年广东省人工智能产业协会青年科技创新奖、2023 年广东省计算机学会科学技术奖科技进步一等奖（第一完成人）、2023 年广东省精准医学科学技术奖科技

创新一等奖（第一完成人）、2023 年广东高校最具转化价值知识产权奖（第一完成人）、2018 年中国计算机学会科学技术奖技术发明二等奖等奖励，获评国际计算机学会（ACM）特色杰出演讲者、中国计算机学会杰出演讲者、中国计算机学会杰出会员等荣誉称号。

特邀讲者简介



特邀讲者 1：张艳宁 西北工业大学副校长

报告题目：以文为媒：高阶语义知识引导的视觉重建、感知与理解

个人简介：张艳宁，西北工业大学教授、博士生导师，教育部“长江学者”特聘教授、中组部首批“万人计划”科技创新领军人才、全国创新争先奖获得者、973 首席科学家，现任西北工业大学副校长。长期从事动态视觉计算理论、技术研究和系统研制，并与航天、航空等方面的国家重大需求相结合，先后主持和承担某基础加强重点项目、973 项目、863、国家自然科学基金重点项目、总装预研等国家级项目 40 余项，带领团队在空间环境感知、空天地海一体化大数据应用技术方面获得了丰富成果。在 IEEE TPAMI、IEEE TIP、IJCV、CVPR、ICCV 等国内外本领域权威期刊和重要国际会议上发表论文 150 余篇，指导学生发表 CVPR 2023 最佳学生论文、ICCV 2011 最佳学生论文，出版专著 4 部，获授权发明专利 150 余项，以第一完成人获国家技术发明二等奖 2 项、国家教学成果二等奖 1 项、国防技术发明一等奖 2 项、陕西省科学技术一等奖 2 项等。



特邀讲者 2: 董晶 中国科学院自动化研究所研究员

报告题目: 数字内容生成与可信鉴别研究浅析

个人简介: 董晶(jdong@nlpr.ia.ac.cn), 中国科学院自动化研究所研究员、博士生导师, CSIG/CCF/IEEE 高级会员, CAAI 杰出会员, 中国科学院青年创新促进会会员, 国家高层次青年人才计划入选者, 目前担任中国图象图形学学会 (CSIG) 理事、副秘书长、女工委秘书长, IEEE 亚太区执委、奖励委员会主席、IEEE 信号处理协会全球成员发展主席。主要从事人工智能安全与对抗、计算机视觉、多媒体内容取证等前沿方向的技术研究, 她先后以课题 (子课题) 负责人承担了 4 项国家自然科学基金、3 项国家重点研发课题、1 项工信部国家重大专项子项及 20 余项省部级科研项目 (课题)。已在国际权威期刊及学术会议上发表学术论文 90 余篇, 已授权 26 项中国专利含 3 项美国专利。她曾获 4 次最佳 (优秀) 论文奖、2016 年度 IBM 学院奖、2019 年度中国人工智能学会杰出贡献奖、2020 年度 CSIG 石青云女科学家奖、2021 年度北京青年优秀科技论文奖、2021 年度 CSIG 科技奖二等奖 (排名第一)、2021 年度吴文俊人工智能科学技术奖 (技术进步科普类)、2022 年度中国发明协会创业创新奖一等奖 (排名第一)、2022 年度国家广电总局 MediaAIAC 大赛深度合成技术应用类一等奖 (排名第一) 以及 2023 年度吴文俊人工智能科学技术奖 (技术发明一等奖、排名第二)。



特邀讲者 3：杨欣 华中科技大学教授

报告题目：视觉空间计算方法及应用

个人简介：杨欣(xinyang2014@hust.edu.cn)，华中科技大学电信学院教授，国家优青获得者，主要从事三维视觉和医学影像分析领域研究工作，已在 TPAMI、IJCV、TMI、CVPR、MICCAI 等国际权威期刊及学术会议上发表学术论文 90 余篇，授权美国/中国发明专利 20 余项，出版英文书籍章节 2 部；获得湖北省技术发明一等奖（排名 1）、中国图象图形学会石青云女科学家奖，CCF B 类会议 ISMAR 最佳论文提名（排名 1）等奖励；入选斯坦福全球 2% 顶尖科学家榜单；指导学生获得全国互联网+金奖和银奖等国内外重要竞赛奖项 10 余次；担任 IEEE-TMI、IEEE-TVCG、Multimedia System 期刊编委及 CVPR、MM 和 MICCAI 领域主席、IEEE BISP 技术委员会委员，CSIG 青工委副秘书长，CCF 多媒体专委会副秘书长等职务；中国图象图形学会杰出会员。

特邀讲者 4: 张铭津 西安电子科技大学教授



报告题目: 跨域推理赋能的智能图像处理

个人简介: 张铭津(mjinzhang@xidian.edu.cn), 西安电子科技大学通信工程学院教授, 博士生导师, 国家级青年人才。2015 至 2016 年赴澳大利亚悉尼科技大学访问研究。长期围绕计算机视觉、人工智能、跨域图像处理开展研究, 成果应用于高分辨对地观测、深空探测、集成电路、智能安防等领域。曾入选中国科协青年人才托举工程、中国图象图形学会石青云女科学家、获吴文俊中国人工智能优秀青年奖、中国图象图形学会优秀博士学位论文, 指导学生获中国“互联网+”大学生创新创业大赛冠军等。迄今为止, 作为第一作者/通讯作者在领域内国际重要期刊及会议发表学术论文 40 余篇, 其中包含 IEEE TIP、TCyb、TNNLS、CVPR、ICCV 等, 担任 TVCJ、Applied Sciences 等 SCI 期刊编委, 担任 CVPR 等国际会议高级程序委员会委员。



特邀讲者 5: 王路 哈尔滨工程大学副教授

报告题目: 海杂波耦合情况下舰船尾迹检测技术

个人简介:王路(wanglu2019@hrbeu.edu.cn), 工学博士(博士后), 副教授, 博士生导师, 入选国家级青年人才, 黑龙江省高层次人才。IEEE/IEICE/CCF 会员、CCF 协同计算专委会执行委员、省人工智能学会理事、省物联网专委会执行委员, 中国图象图形学学会会员。曾获 CSC/日本政府国费(MEXT) 奖学金赴日攻读博士学位, 毕业于日本庆应义塾大学。曾在大阪大学担任研究助理和庆应义塾大学担任助理教授职位。研究领域涉及智能信息感知、跟踪、数据挖掘等。近年来, 发表 SCI/EI 等检索学术论文 46 篇, 第一作者论文 33 篇, 在国际会议上作学术报告超过 24 次。作为项目负责人主持科研项目 11 项, 参与科研项目 6 项。2019 年获得日本电子情报通信学会优秀科研奖、2023 年国际会议 IoTCIT 优秀论文奖等。此外, 作为国际会议主席组织了 MSN 2020 Workshop、2023IoTCLT Workshop、2023ChineseCSCW 竞赛等。担任 AJRS、CST 期刊编委, IEEE Globecom 2023、2021, IEEE ICC 2021、2020, WCNC 2021 等会议的程序委员会委员, 作为审稿人服务于 IEEE TNSE, IEEE TVT, IEEE TGRS, IEEE WCL 等期刊。

专题论坛：三维重建与生成

(论坛主席：陶文兵，华中科技大学人工智能与自动化学院教授，

郭裕兰，国防科技大学电子科学学院副教授)

(10月18日 17:30-19:30，地点：多功能厅)

时间	报告嘉宾姓名、单位、及联系信息	报告题目
10月18日		
17:30-18:00	申抒含 中国科学院自动化研究所 shshen@nlpr.ia.ac.cn	大规模场景高精度位姿解算与结构化三维建模
18:00-18:30	高林 中国科学院计算技术研究所 gaolin@ict.ac.cn	面向高真实感建模的三维重建与生成方法研究
18:30-19:00	彭思达 浙江大学 pengsida@zju.edu.cn	面向物理世界仿真的三维重建与生成
19:00-19:30	惠乐 西北工业大学 huile@nwpu.edu.cn	基于3D几何感知的可变形高斯溅射新视角合成

论坛介绍

三维重建与生成是三维视觉领域的重要研究方向。三维内容在展示物体几何结构、表现物体运动、呈现人机交互等领域具有二维内容不可比拟的优势。近年来，随着三维视觉技术的快速发展，三维重建与生成的相关成果已广泛应用于无人驾驶、机器人、数字人等工业与民生领域，且具有巨大的发展潜力。本次论坛的目的在于汇聚计算机视觉、计算机图形学、机器人学、模式识别等领域从事三维重建与生成的顶尖学者，分享该领域的最新技术进展，探讨三维重建与生成在不同领域的落地应用，剖析当前研究面临的困难、挑战和潜在机遇，为国内外从事相关研究的专家学者与从业人员创建一个交流与合作的平台。

论坛主席



个人简介：陶文兵，华中科技大学人工智能与自动化学院教授/博士生导师。曾连续 6 年入选中国高被引学者（Most Cited Chinese Researchers）计算机领域榜单，发表各类学术论文 150 余篇，其中以第一作者和通讯作者在 IEEE 汇刊等中科院二区以上国际权威期刊及计算机学会 A 类国际顶级会议上发表论文 60 余篇，所发表的论文引用近 6000 次。多次应邀在新加坡南洋理工大学、挪威奥斯陆大学、香港浸会大学、澳门大学等进行合作研究并担任研究员职位。主持和参与完成国家自然科学基金重大项目、重点项目、面上项目，以及 863 计划、十二五预研、十三五预研等人工智能、计算机视觉和图像处理等研究领域相关课题 30 余项。授权国家发明专利 38 项。



个人简介：郭裕兰，国防科技大学电子科学学院副教授。主要研究领域为三维视觉，包括三维重建、点云语义理解及点云压缩编码等。在 IEEE TPAMI 和 CVPR 等期刊和会议发表学术论文 200 余篇，谷歌学术总被引 1.5 万余次，连续四年入选 Elsevier 中国高被引学者，连续 2 年入选全球前 2% 顶尖科学家榜单、百度全球华人 AI 青年学者，获吴文俊人工智能优秀青年奖、ACM China SIGAI Rising Star。担任中国图象图形学学会三维视觉专委会秘书长，IEEE Transactions on Image Processing、Computers & Graphics、Visual Computer 等国际学术期刊编委，CVPR 2023/2021、ICCV 2021、ECCV 2024、NeurIPS 2024、ACM Multimedia 2021 等国际会议领域主席。

特邀讲者简介



特邀讲者 1：申抒含 中国科学院大学教授

报告题目：大规模场景高精度位姿解算与结构化三维建模

报告摘要：对大规模复杂场景进行高精高效的三维重建与理解，包括三维几何重建、三维语义理解、三维结构表达等，是数字城市、无人系统、工业孪生等领域的共性需求。在这一研究中，计算机视觉、摄影测量、机器人等多个领域的研究不断交叉融合，几何视觉方法与深度学习方法也在不断发生碰撞。本次报告将从几何和学习的视角介绍我们在大规模场景高精度位姿解算与结构化三维建模方面的最新研究进展，以及在实景三维重建、室内 BIM 建模、工业三维孪生等领域的实际应用。

个人简介：申抒含，中国科学院自动化研究所研究员、博士生导师，中国科学院大学人工智能学院岗位教授，中国科学院工业视觉智能装备技术工程实验室副主任。研究领域为三维计算机视觉理论与应用，包括大规模场景三维重建、智能机器人三维环境感知、场景三维语义理解等。在计算机视觉、摄影测量、机器人等领域国际期刊和国际会议，如 IEEE Trans、IJCV、ISPRS JPRS、CVPR、ICCV、ICRA、IROS 等发表论文 90 余篇。作为项目负责人主持国家自然科学基金联合重点项目、北京市自然科学基金联合重点项目、以及各类企业研发项目 20 余项。曾获 2016 年 ACM 北京新星奖、2018 年中国图像图形学会科学技术二等奖、2023 年中国测绘学会科学技术一等奖、2023 年中国自动化学会自然科学一等奖等。

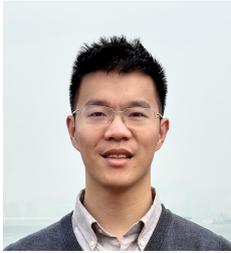


特邀讲者 2: 高林 中国科学院大学教授

报告题目: 面向高真实感建模的三维重建与生成方法研究

报告摘要: 近期, 生成式人工智能取得了快速的发展, 诞生了诸如 GhatGPT, StableDiffusion, 可灵等生产力工具, 正在成为构建新质生产力的关键要素。其中三维模型是非常普遍和被广泛应用的数字媒体内容, 高效的生成高质量的三维模型是学术界和工业界所关注的前沿课题。生成式人工智能的发展为上述课题的研究带来了新的研究思路和研究方法, 并形成了系列基于生成式人工智能的三维模型建模方法 (3D AIGC)。在本次报告中, 我们将分享面向高真实感建模的三维重建与生成的背景, 现状与相应的技术挑战, 并将探索如何重建和生成高质量的三维模型。包括重建具有复杂拓扑结构和材质的三维模型以及通过线稿几何交互的方法生成高质量的三维模型, 并将介绍最新的几何表征技术的发展, 包括神经辐射场, 高斯泼溅方法和高斯网。

个人简介: 高林, 中国科学院计算技术研究所泛在计算系统研究中心副主任、研究员、博士生导师、中国科学院大学岗位教授。研究方向为计算机图形学、三维计算机视觉。在 SIGGRAPH、TPAMI、TVCG 等期刊会议发表论文 100 余篇, 研发的人脸 AIGC 的 APP 被全球 180 多个国家或者地区的用户所使用。现任或者曾任 GDC 2024 大会联合程序主席, SGP 2023 大会联合主席, China 3DV 2023 程序委员会联合主席, SIGGRAPH 2023-2024 技术论文程序委员会委员, NeurIPS 2024 领域主席, IEEE TVCG 编委, CSIG 智能图形专委秘书长, 入选国家自然科学基金委优青, 北京市杰青, 英国皇家学会牛顿高级学者, 曾获得亚洲图形学会青年学者奖, 吴文俊人工智能优秀青年奖, CCF 技术发明一等奖, CCF CAD&CG 开源软件奖等奖励。



特邀讲者 3：彭思达 浙江大学研究员

报告题目：面向物理世界仿真的三维重建与生成

报告摘要：物理世界仿真是对人们日常生活进行真实、立体、时序化反映和表达的数字空间。传统多视图重建算法是构建数字三维世界的重要工具。然而，传统重建算法依赖精细的数据采集过程，并且难以处理复杂动态场景。针对这些挑战，讲者基于多视图几何和先验学习等思想，在场景重建以及物体生成三方面提出了新技术，提升了场景仿真的质量与效率。

个人简介：彭思达，浙江大学软件学院“百人计划”研究员，博士生导师，启真优秀青年学者。2023 年获得浙江大学计算机科学与技术博士学位。研究方向为三维计算机视觉，代表工作为 EasyVolCap、Neural Body、PVNet。至今在 TPAMI、CVPR、ICCV 等期刊或会议发表四十余篇论文，谷歌学术引用 3900 余次，其中一篇一作论文获得 CVPR 最佳论文提名，在 GitHub 开源获得上万次 stars。曾获得 2023 年全球 IMC 三维重建挑战赛冠军、2023 年世界人工智能大会云帆奖-明日之星、2022 Apple Scholar、2020 年 CCF-CV 学术新锐奖、2023/2021 年中国 CCF 图形开源软件奖。



特邀讲者 4: 惠乐, 西北工业大学大学副教授

报告题目: 基于 3D 几何感知的可变形高斯溅射新视角合成

报告摘要: 针对新视角合成任务, 现有基于神经辐射场 (NeRF) 的解决方案通过隐式方式学习形变, 难以结合 3D 场景的几何信息, 导致学习到的形变可能与几何不一致, 从而影响动态视图合成和 3D 动态重建的效果。3D 高斯溅射提供了一种全新的 3D 场景表示方式, 在此基础上我们考虑利用 3D 几何信息来帮助学习复杂的 3D 形变。通过场景表示为 3D 高斯分布的集合, 每个 3D 高斯分布都经过优化, 能够随时间移动、旋转和进行尺度变化, 精准建模形变过程。为在形变过程中引入 3D 场景几何约束, 我们显式提取 3D 几何特征并将其融入 3D 形变的学习中。此外, 我们建议使用连续的 6D 旋转表示用于表示旋转变化。在合成和真实数据集上的广泛实验结果证明了提出方法的优越性。

个人简介: 惠乐, 西北工业大学大学副教授, 入选香江学者计划、西北工业大学翱翔新星, 主持国家自然科学基金青年项目, 获南京理工大学优秀博士论文。惠乐研究方向聚焦于三维视觉, 主要包括三维场景语义与实例分割、三维目标跟踪、三维点云生成、三维重建等, 主要成果发表在顶级期刊和会议 TIP、CVPR、ICCV、ECCV、NeurIPS 等共 20 篇论文, 其中 10 篇一作论文。长期担任国际人工智能顶刊和顶会 TPAMI、CVPR、ICCV、ECCV、NeurIPS 等的审稿人。

专题论坛：情智兼备数字人与机器人的研究

（论坛主席：赵思成，清华大学副研究员，

高跃，清华大学长聘副教授，

杨巨峰，南开大学计算机学院教授，

姚鸿勋，哈尔滨工业大学计算学部长聘教授）

（10月18日 17:30-19:30，地点：学术厅）

时间	报告嘉宾姓名、单位、及联系信息	报告题目
10月18日		
17:30-17:50	胡斌 北京理工大学 bh@bit.edu.cn	心智工作机制与精神障碍诊疗
17:45-18:00	吕宝粮 上海交通大学 blu@cs.sjtu.edu.cn	情感智能与情感脑机接口：现状与挑战
17:50-18:10	刘青山 南京邮电大学 qslu@nuist.edu.cn	多模态视觉情感计算
18:10-18:30	刘洪海 哈尔滨工业大学(深圳) honghai.liu@hit.edu.cn	情智兼备的孤独症早期诊疗
18:30-18:50	郑文明 东南大学 wenming_zheng@seu.edu.cn	多模态情感人机交互与生理心理信息感知
18:45-19:30	所有报告嘉宾与组织者	Panel 讨论

论坛介绍

近日，中国科协发布 2024 重大科学问题、工程技术难题和产业技术问题，其中由中国图象图形学学会情感计算与理解专委会提出的“情智兼备数字人与机器人的研究”成功入选 2024 重大科学问题。情感是人类内在的情绪、感觉和情感状态，涵盖了广泛的情绪体验。作为人类生活中不可或缺的一部分，情感影响着个体的认知、行为和决策，也在人际交往和社会关联中扮演着重要角色。情感计算是一门跨学科交叉的研究领域，涵盖了计算机科学、心理学、认知科学、神经科学等多个学科的知识和技术，旨在理解、模拟和处理人类情感。情感计算的主要研究目标是深入研究情感与认知之间的相互作用，通过探索情感如何影响认知过程，以及认知如何调节情感体验，从而更全面地理解人类情感认知的本质。同时，不断优化情感识别和表达技术，从语音、文本、图像等多种数据源中准确识别和表达情感，更好地将情感信息整合到计算机界面和智能体中，以实现更准确、更自然的情感交流。在此基础上，开发出具有感知、处理和交互功能的情智兼备的通用智能体，不仅能够识别和生成情感内容，还能够理解情感背后的认知过程，并相应地调整其行为和反应。

数字人和机器人是人工智能众多前沿技术的集成者，融合了机器学习、自然语言处理、感知技术等多个子领域的成果。数字人是指通过计算机软件模拟出的具有人类特征和行为的虚拟实体，具有自己的思维、情感和个性，与真实世界中的人类进行交互，并扮演各种角色。机器人则是在此基础上构建而成的，包含机械结构、传感器、执行器和控制系统等多个部分的智能化系统。现有研究成果主要集中在感知、决策、学习上，在此基础上引入人类情感，构建情智兼备的数字人和机器人，将使人机交互更加自然和人性化，有利于机器和数字实体更好地适应人类社会的需求，实现更加智能和人性化的交互体验，对推动通用人工智能技术的发展具有重要意义，将在医疗、教育、工业等领域展示出广阔的应用前景，有效服务民生改善、国防建设等国家重大需求。当情智兼备的数字人与机器人的研究仍面临情感产生机理不明、多模态情感智能解译难、个体间情感差异大和仿生化情感交互难等挑战。该论坛将邀请国际知名专家共同探讨该领域的前沿进展，包括专题报告与 panel 讨论，为学术界和产业界相关研究和技术人员提供构造情智兼备数字人与机器人的新思路和新方法。

论坛主席



个人简介：赵思成，清华大学副研究员，国家级青年人才，ACM/IEEE/CCF/CSIG 高级会员，哈尔滨工业大学博士，加州大学伯克利分校和哥伦比亚大学博士后。研究方向为情感计算、多媒体、机器学习等。发表 IEEE/ACM 汇刊、CCF 推荐 A 类论文 60 余篇，谷歌学术引用 10000 余次，H 指数为 50。担任中国图象图形学学会 CSIG 情感计算与理解专委会秘书长、IEEE TIP/IEEE TAFPC 等国际期刊编委、IEEE TFS/ACM TOMM 等国际期刊首席客座编辑、2023 第一届 CSIG 情感智能大会程序主席、NeurIPS/ACM MM/ECCV/AAAI/IJCAI/ICASSP 等国际会议领域主席或资深程序委员。主持或参与国家自然科学基金、国家重点研发计划、CCF-滴滴盖亚学者科研基金等项目。获得 CSIG 优秀博士论文、ACM SIGMM 中国新星奖等奖励，入选 AI 2000 人工智能全球最具影响力学者榜单和全球前 2% 顶尖科学家榜单。



个人简介：高跃，清华学长聘副教授、博士生导师，国家青年特聘专家、DeepTech 2022 年中国智能计算科技创新人物。主要研究领域为人工智能、计算机视觉及医学图像处理，提出并发展了超图计算理论及方法，形成了面向有效数据不足、信息关联复杂等环境的新方法。研究成果发表论文百余篇，论文引用万余次，出版《Hypergraph Computation》等英文专著，多次入选科睿唯安高被引科学家和爱思唯尔中国高被引学者，担任 International Journal of Computer Vision 和 Medical Image Analysis 等国际重要期刊编委。曾获得中国电子学会自然科学一等奖、福建省科技进步一等奖等。



个人简介：杨巨峰，南开大学计算机学院教授、博士生导师，数据与智能系统安全教育部重点实验室副主任、天津市媒体计算技术工程研究中心副主任，国家“万人计划”青年拔尖人才、天津杰青、南开百青。研究方向是计算机视觉、机器学习、多媒体计算，发表 IEEE/PAMI/CVPR 等高水平学术论文 60 余篇。任中国计算机学会计算机视觉专委会副秘书长，中国图象图形学学会情感计算与理解专委会常务委员，中国中文信息学会情感计算专委会常务委员，IEEE 情感计算汇刊（TAC）编委，中国计算机视觉大会（CCCV 2017）组织主席，视觉与学习青年研讨会（Valse 2022）Workshop 主席，国际计算机视觉与模式识别大会（CVPR 2024）领域主席。研究成果获第十届吴文俊人工智能自然科学二等奖。



个人简介:姚鸿勋，哈尔滨工业大学计算学部长聘教授，黑龙江省人民政府特殊津贴专家，教育部“新世纪优秀人才”，中国图象图形学学会情感计算与理解专业委员会主任。主要研究领域为计算机视觉智能、多媒体数据分析与理解、情感计算等。已发表 ICCV, CVPR, ACM MM 等顶级国际会议及 IJCV, TPAMI, TIP, TMM 等高影响因子国际期刊文章学术论文 300 余篇，H 指数>50，Google 引用量 12000+，入选全球人工智能 TOP 2000 学者榜单。主持国家自然科学基金重点项目、新一代人工智能国家科技重大专项课题，获国家及省部级自然科学奖项 4 项，获中国计算机专业优秀教师奖(2021)，已培养学生成为“国家级高层次人才”、“国家级青年人才”等 6 名。

特邀讲者简介



特邀讲者 1：胡斌 北京理工大学

报告题目：“心智工作机制与精神障碍诊疗”

报告摘要：心智是脑和神经的功能。心智工作机制是人脑接受外界信息，经过加工处理，转换成内在的心理活动，从而获取知识或应用知识的过程，是人类心理与智能的表现。心智工作机制不清将会严重阻碍脑机制、类脑计算和脑疾病的深入研究。在脑功能性疾病的研究中，由于心智工作机制不清，抑郁等临床诊断方法缺少多维度、动态、持续性客观量化指标，以症状描述（如访谈、量表等）为主，主观性大、误诊率高；同时，在治疗方法上难以实现个性化的治疗，治疗有效率低。利用医学电子、人工智能等信息技术揭示脑功能性疾病的心智工作机制，将从根本上促进其临床诊断方法由症状描述型向数据驱动型的转变，实现对脑功能性疾病的精细化分层分类；继而依据不同疾病的发生发展规律，实施个性化的非药物或药物治疗，实现早诊早治，提高诊疗有效率。

个人简介：胡斌，国家海外高层次人才入选者，973 首席科学家，IEEE/IET/AAIA Fellow，教育部计算机学科教指委委员，科技委委员，IEEE Transaction on Computational Social Systems 主编，国际社会神经科学中国分会副主席等；获 2019 年度中国专利金奖，2018 年度国家技术发明奖二等奖、2016 年度教育部技术发明奖一等奖、2022 年度甘肃省自然科学特等奖，2023 年度全国创新争先奖，（均为第一完成人）等，2020 年全国先进工作者，入选 2020-2023 年高被引科学家名单。



特邀讲者 2：刘青山 南京邮电大学教授

报告题目：“多模态视觉情感计算”

报告摘要：情感计算的目标是使计算机能自动识别情感并与表达情感，从而建立适应人类情感的和谐人机环境，也是人工智能领域中的一个热点研究方向。基于表情、肢体动作等视觉信息是情感表达的重要载体。本报告将简单汇报一下研究背景、现状，并重点汇报团队在动态表情特征学习、姿态估计、及其结合语音语言的多模态视觉情感计算等方面的研究进展。

个人简介：刘青山，南京邮电大学教授，主要从事模式识别、图像理解，人工智能+交叉研究等，近年来主持承担了国家杰出青年基金项目、科技部 2030 人工智能重大专项项目、国家自然科学基金重点项目等。先后入选教育部特聘教授、科技部中青年创新领军人才、江苏省双创领军人才、江苏省优秀教育工作者称号、江苏省特聘教授等。曾获江苏省科学技术一等奖、教育部自然科学二等奖、中国电子学会自然科学一等奖等。现兼任中国计算机学会计算机视觉专委会副主任、中国图形图像学会学术工委主任、江苏省人工智能学会副理事长等。



特邀讲者 3：刘洪海 哈尔滨工业大学（深圳）教授

报告题目：“情智兼备的孤独症早期诊疗”

报告摘要：孤独症早期筛查与干预是缓解孤独症疾病最有效的策略与技术。讲者汇报了孤独症诊疗现状和其团队研发的孤独症诊疗技术与设备，围绕“动作模仿、共同注意和角色变换”典型行为，诱发孤独症受试者视线等共情行为表征，推荐知识图谱驱动个性化干预，并研讨情智兼备智能诊疗系统前沿技术与挑战。

个人简介：刘洪海，哈尔滨工业大学（深圳）教授、欧洲科学院院士、国际电子电气工程师协会会士（FIEEE）、英国技术工程院会士（FIET）。长期从事人机交互、具身智能、医疗机器辅助系统理论及脑疾病应用等方面的研究，主持国家重点研发计划、国家自然科学基金重点等项目，研究成果已在多自由度灵巧假肢、自闭症早期诊疗和脑卒中精准诊疗等领域取得了一系列具有国际影响力的创新成果，并得到成功应用。已发表 400 多篇国际权威杂志和会议论文，现担任《IEEE Trans. Industrial Informatics》共同主编、《IEEE Trans. Cybernetics》等期刊编委。



特邀讲者 4：郑文明 东南大学教授

报告题目：“多模态情感人机交互与生理心理信息感知”

报告摘要：机器情感智能是智能机器人研究的重要目标，以表情、语音、脑电为代表的多模态生理心理信息感知技术是实现机器情感智能的关键。报告围绕多模态情感人机交互的应用需求，简要介绍以脑电、表情、语音为代表的生理心理信息感知技术及应用。

个人简介：郑文明，东南大学二级教授、博士生导师，国家级领军人才，IET Fellow，儿童发展与学习科学教育部重点实验室主任，中国认知科学学会理事。主要从事情感计算、计算机视觉，模式识别等领域研究。担任 IEEE Transactions on Affective Computing 等期刊编委，研究成果获国家技术发明二等奖 1 项、省部级二等奖 4 项、IEEE 计算机协会最佳论文奖、IEEE 情感计算汇刊最佳论文奖。

专题论坛：多模态连续学习论坛

(论坛主席：尹宝才，北京工业大学教授，

王博岳，北京工业大学副教授，

胡永利，北京工业大学教授，

高源，武汉大学副教授)

(10月18日 17:30-19:30，地点：塔里木厅)

时间 10月18日	报告嘉宾姓名、单位、及联系信息	报告题目
17:30-17:50	邓成 西安电子科技大学 chdeng@mail.xidian.edu.cn	在线增量学习
17:50-18:10	魏云超 北京交通大学 yunchao.wei@bjtu.edu.cn	Continual Learning Meets Real-World Visual Perception
18:10-18:30	赵鑫 中国人民大学 Xin Zhao batmanfly@gmail.com	面向大模型训练的数据工程方法
18:30-18:50	崔振 南京理工大学 zhen.cui@njust.edu.cn	小样本增强式持续学习研究
18:50-19:10	赵洲 浙江大学 zhaozhou@zju.edu.cn	持续学习视角下的 多模态统一离散化表征
18:50-19:10	刘夏雷 南开大学 xialei@nankai.edu.cn	基于图文预训练模型的连续学习方法研究
19:10-19:30	高源 武汉大学 ethan.y.gao@gmail.com	General-Purpose Multi-Task Learning with Adaptive Task-Knowledge Sharing

论坛介绍

多模态连续学习模仿了人类学习过程中的多感官感知和记忆保持功能，旨在提高模型的泛化性、灵活性和实用性，近些年吸引了来自计算机视觉、自然语言处理、脑科学等多个学科领域学者们的广泛关注。如何突破跨媒体数据的异构鸿沟和语义鸿沟、如何避免在学习新任务时遗忘旧知识造成的灾难性遗忘、如何自适应建模任务间的内在联系以通用地适配任意输入任务并提升模型泛化能力、以及开展多模态大模型的连续学习必要性，都是多模态连续学习研究的关键和热点问题。为深入探讨这些挑战，本论坛汇集相关领域的专家学者共同探索多模态连续学习的最新研究进展与面临的挑战，并对其未来发展趋势进行交流和讨论。期待通过这次论坛，能够激发新的思考，推动这一研究的创新和应用。

论坛主席



个人简介：尹宝才(ybc@bjut.edu.cn)，博士生导师，北京学者，国家杰出青年基金获得者，北京人工智能研究院院长，多媒体与智能软件技术北京市重点实验室主任，ACM北京分会副主席。主要研究领域包括多媒体技术、计算机视觉、模式识别和人工智能。近五年，作为课题负责人承担国家自然科学基金重大项目课题2项，作为负责人承担国家自然科学基金重点项目1项。在图像视频等高维数据压缩编码、复杂场景对象建模与理解、交通大数据分析等方面取得了一系列创新成果，发表IEEE/ACM汇刊或CCF-A类会议学术论文60余篇，授权相关专利60余项，制定国家和行业标准10余项。2009年获“享受国务院特殊津贴专家”，2014年主持的交通视频处理团队入选“北京市学术创新团队”，2021年入选“北京学者”。获得多项科学技术奖励，2003年“基于多功能感知理论的中国手语识别与合成研究”成果获得国家科技进步二等奖；2007年“基于WEB的手语播报系统与手语普适终端”成果获教育部高等学校科技进步一等奖；2020年“多源高维数据协同表征及应用”成果获得中国人工智能学会吴文俊人工智能科技进步一等奖。



个人简介：王博岳(wby@bjut.edu.cn)，北京工业大学副教授、博士生导师。研究方向为跨模态分析，以第一/通讯作者发表IEEE/ACM汇刊、CCF-A会议论文20余篇；主持科研项目10余项，主要包括国家重点研发计划课题、国家自然科学基金重点研究计划培育项目/面上项目/青年项目、北京市自然科学基金青年项目等；中国图象图形学学会多媒体专委会委员、中国图学学会可视化与认知计算专业委员会委员、北京人工智能学会理事；获“中国图学学会优秀博士学位论文奖”，入选“北京工业大学高端人才队伍建设计划”、“北京工业大学新锐青年学者”，入选北京市科协“青年人才托举工程”。



个人简介：胡永利(huyongli@bjut.edu.cn)，北京工业大学教授，博士生导师，北京人工智能研究院研究员。北京市百千万人才工程入选者，北京市高层次创新创业人才支持计划领军人才。长期从事模式识别、计算机视觉、机器学习和智能交通等方面的研究，近年来关注图像视频聚类、多视聚类、跨媒体和图网络等方面的研究。主持国家自然科学基金联合基金重点、面上和北京市科技计划等项目 10 余项。共发表论文 110 余篇，其中包括 IEEE TIP、TNNLS、TCYB、TMM、TCSVT、TITS、ACM TKDD 等 IEEE/ACM 汇刊和 CVPR、AAAI、IJCAI、ACM MM 等 CCF A 类会议论文 30 余篇。获得国家发明专利授权 22 项。获 2020 吴文俊人工智能科技进步奖一等奖、2020 青岛市科技进步奖二等奖、2017 年和 2018 年国家自然科学基金委水下机器人目标抓取大赛二等奖。



个人简介：高源(ethan.y.gao@gmail.com)，武汉大学副教授、博士生导师；曾任腾讯 AI Lab 高级研究员、拼多多视觉算法部算法专家及研究主管；研究方向为多模态多任务学习、高效深度学习、三维计算机视觉；在包括 Nat. Comm., IJCV, TIP, ICML, ICLR, CVPR, ECCV 等顶级期刊和会议上发表论文 30 余篇；申请授权发明专利 16 项，含国际专利 8 项；主持研发的“人脸识别与活体验证系统”在腾讯云部署、“轻量化通用多任务图像理解”支持拼多多算法中台，服务于多家政府和金融机构，支持了多个千万级规模的消费者应用，取得了行业领先效果。

特邀讲者简介

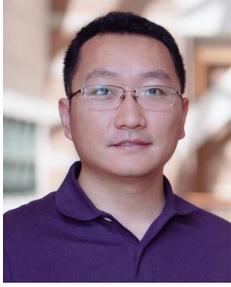


特邀讲者 1：邓成 西安电子科技大学教授

报告题目：在线增量学习

报告摘要：人类与生俱来具有终身不断获取、整合和迁移知识的能力，这种学习能力被称之为增量学习（Incremental Learning）。在机器学习领域，特别是深度学习模型提出以来，增量学习致力于解决模型训练的一个普遍缺陷——灾难性遗忘（Catastrophic Forgetting），即在新任务上训练时，在旧任务上的表现通常会显著下降。本报告以增量学习的两个范式——任务增量学习和类增量学习为主线，首先探讨分析近年来的研究进展，并介绍本组在该领域取得的最近研究成果。

个人简介：邓成(chdeng@mail.xidian.edu.cn)，西安电子科技大学二级教授，博士生导师。国家级高层次人才，国家百千万人才工程入选者，国家有突出贡献中青年专家，陕西省重点科技创新团队负责人，陕西省教学名师，陕西省师德标兵，教育部电子信息类教学指导委员会秘书长。长期从事人工智能领域的研究工作，主持国家自然科学基金重点项目、科技部重点研发计划等项目近 30 项。在本领域国际一流期刊和 CCF A 类会议上发表论文 200 余篇，连续多年入选爱思唯尔中国高被引学者榜单。研究成果获 2019 年、2023 年陕西省自然科学一等奖（第 1）、2016 年国家自然科学基金二等奖（第 3）。



特邀讲者 2: 魏云超 北京交通大学教授

报告题目: Continual Learning Meets Real-World Visual Perception

报告摘要: 连续学习下的视觉内容感知研究聚焦于如何使模型不断适应新环境、识别新类别最终达成模型自身认知能力的提升。在本次报告中,魏云超教授将围绕如何基于预训练模型开展连续学习?如何解决精细化视觉感知任务中背景漂移问题?在多模态大模型/生成式模型时代开展连续学习的必要性?等问题展开讨论并介绍其团队在执行科技创新 2030-新一代人工智能重大项目“连续学习的理论和方法”过程中产生的相关研究成果。

个人简介: 魏云超(yunchao.wei@bjtu.edu.cn), 北京交通大学计算机学院教授、副院长, 国家级讲席教授。曾在新加坡国立大学、美国伊利诺伊大学厄巴纳-香槟分校、悉尼科技大学从事研究工作。入选 MIT TR35 China、百度全球高潜力华人青年学者、《澳大利亚人》TOP 40 Rising Star; 获世界互联网大会领先科技奖(2023)、教育部高等学校自然科学奖一等奖(2022)、中国图象图形学学会科技技术奖一等奖(2019)、澳大利亚研究委员会青年研究奖(2019)、IBM C3SR 最佳研究奖(2019)、计算机视觉世界杯 ImageNet 目标检测冠军(2014)及多项 CVPR 竞赛冠军; 发表 TPAMI、CVPR 等顶级期刊/会议论文 100 多篇, Google 引用超 22000 次。目前主要研究方向包括面向非完美数据的视觉感知、多模态数据分析、生成式人工智能等。



特邀讲者 3: 赵鑫 中国人民大学教授

报告题目: 面向大模型训练的数据工程方法

报告摘要: 近年来,大语言模型的研发受到了广泛关注,一般包括预训练与后训练两个阶段,其中预训练阶段是建立模型能力的关键。本次报告将聚焦预训练与继续预训练过程的基本技术路线,将针对其中涉及的数据工程方法展开讨论。除了介绍如何高效利用已有数据外,还将介绍大规模合成数据在继续预训练中的应用方法。特别地,我们将以 Llama 3 为例,完整介绍继续预训练所涉及的数据工程方法。

个人简介: 赵鑫(batmanfly@gmail.com),中国人民大学高瓴人工智能学院教授。2014年7月于北京大学获得博士学位,随后进入中国人民大学工作至今。研究领域为信息检索与自然语言处理,共计发表论文200余篇,谷歌学术引用2.2万余次,曾主导研发了玉兰大语言模型,组织编写了大语言模型综述论文《A Survey of Large Language Models》(预印版文章)以及《大语言模型》中文书。曾荣获2020年吴文俊人工智能优秀青年奖、ECIR 2021 时间检验奖,CCF-IEEE CS 青年科学家奖。

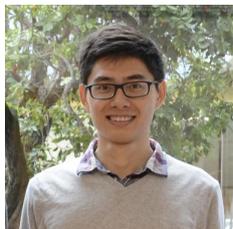


特邀讲者 4: 崔振 南京理工大学教授

报告题目: 小样本增强式持续学习研究

报告摘要: 小样本条件下的持续学习是当前人工智能的重要研究方向。本报告将介绍近期小样本增强式持续学习的研究进展, 并分享我们在这一领域的一些探索性工作。这些研究涵盖样本扩散生成、数据强化精炼和增量微分建模等技术, 旨在帮助大家更深入地了解持续学习的研究动态。

个人简介: 崔振(zhen.cui@njust.edu.cn), 南京理工大学教授、博导。研究方向为计算机视觉和模式识别, 聚焦于视觉计算及图建模学习研究。在 IEEE 汇刊和 CCF A 类会议发表论文 90 余篇。获中国科学院优秀博士论文奖、IEEE Computer Society 最佳论文奖、IEEE T-AFFC 最佳论文奖, 航天集团科技进步二等奖, 多次国内外重大/重要赛事冠亚军等; 入选国家高层次青年人才计划, 承担江苏省杰出青年基金、国家自然科学基金、装发预研等多项项目。



特邀讲者 5: 刘夏雷 南开大学副教授

报告题目: 基于图文预训练模型的连续学习方法研究

报告摘要: 连续学习是新一代人工智能系统的关键技能之一,旨在使系统能够在动态变化的环境中不断学习新知识,同时避免灾难性地遗忘旧知识,以模拟人类学习的方式。本报告将以图文预训练模型为基础,探索多模态模型的连续学习,分别从判别性模型(如 CLIP)角度和生成性模型(如 mini-GPT4)角度,解决目前连续学习所面临的挑战。最后对连续学习领域进展进行总结和展望。

个人简介: 刘夏雷(xialei@nankai.edu.cn),南开大学计算机学院副教授,研究方向为开放环境视觉连续学习。入选南开大学“百名青年学科带头人培养计划”,入选第九届中国科协青年托举计划,博士生导师。博士毕业于西班牙巴塞罗那自治大学,博士后工作于英国爱丁堡大学。长期从事连续学习、无监督学习和小样本学习等面向开放环境的机器学习和计算机视觉问题。至今共发表学术论文 40 余篇,谷歌学术引用 3500 余次。包含国际顶级期刊和会议 TPAMI、NeurIPS、CVPR、ICCV 等,一篇文章入选 CVPR 2022

Best Paper Finalists。担任 VALSE 2022-2024 组委会成员，组织 CVPR 2023 年连续学习 Workshop，获第二届粤港澳大湾区（黄埔）国际算法算例大赛“序列任务的连续学习”冠军。



特邀讲者 6: 高源 武汉大学副教授

报告题目： General-Purpose Multi-Task Learning with Adaptive Task-Knowledge Sharing

报告摘要： 现实世界中大多问题实质上是复杂的系统问题，涉及多个任务的协作。通用多任务学习通过自适应建模任务间的内在联系，以通用地适配任意输入任务，并提升模型泛化能力。本报告围绕通用多任务学习中最本质的负迁移问题展开，结合实际应用中普遍存在的关键限制（如稀疏的标注数据、有限的部署资源、任务间相关度的可解释性），将通用多任务学习建模为任务间自适应特征融合框架，深入探讨如何进行特征融合、在何处进行特征融合、何时进行特征融合、以及哪些任务应进行特征融合等通用多任务学习关键问题，并介绍课题组最新的相关研究进展。

个人简介： 高源(ethan.y.gao@gmail.com)，武汉大学副教授、博士生导师；曾任腾讯 AI Lab 高级研究员、拼多多视觉算法部算法专家及研究主管；研究方向为多模态多任务学

习、高效深度学习、三维计算机视觉；在包括 Nat. Comm., IJCV, TIP, ICML, ICLR, CVPR, ECCV 等顶级期刊和会议上发表论文 30 余篇；申请授权发明专利 16 项，含国际专利 8 项；主持研发的“人脸识别与活体验证系统”在腾讯云部署、“轻量化通用多任务图像理解”支持拼多多算法中台，服务于多家政府和金融机构，支持了多个千万级规模的消费者应用，取得了行业领先效果。

专题论坛：机器视觉与学习论坛

（论坛主席：林宙辰，北京大学教授，

王奕森，北京大学助理教授）

（10 月 19 日 17:30-19:30，地点：多功能厅）

时间	报告嘉宾姓名、单位、及联系信息	报告题目
10 月 19 日		
17:30-18:00	尤著宏 西北工业大学 zhuhongyou@nwpu.edu.cn	生物医学知识图谱表示学习：在分子相互作用预测中的应用
18:00-18:30	李建欣 北京航空航天大学 lijx@buaa.edu.cn	大数据关联建模及长序预测方法
18:30-19:00	李霄剑 合肥工业大学 lixj90@hfut.edu.cn	三维感知驱动的微创手术智能化技术研究
19:00-19:30	江劭玮 杭州电子科技大学 jiangsw@hdu.edu.cn	高通量叠层显微成像技术与应用

论坛介绍

机器学习是一门从已知数据中探寻未知规律和模式的艺术，如何以更少的代价取得更好的学习效果始终是其亟待解决的核心问题。随着机器学习技术在视觉等领域的发展和应用，传统学习算法在逐步向前迭代的同时，各式深度学习方法也在不断推陈出新，本次 Workshop 我们将邀请领域专家作为讲者，分享他们在该领域的最新工作，总结机器视觉与学习领域在相关核心问题上的近期进展和展望未来值得探索的方向。

论坛主席



个人简介：林宙辰，北京大学教授，IAPR/IEEE/CSIG Fellow，国家杰青，中国图象图形学学会机器视觉专委会主任，中国自动化学会模式识别与机器智能专委会副主任。研究领域为机器学习、数值优化。发表论文 230 余篇，英文专著 2 本。获 2020 年度 CCF 科学技术奖自然科学一等奖。多次担任机器学习与人工智能顶级会议 CVPR、ICCV、NIPS/NeurIPS、ICML、IJCAI、AAAI 和 ICLR 的领域主席，曾任 IEEE T. Pattern Analysis and Machine Intelligence 编委，现任 ICPR 2022 程序共同主席、ICML 2022 资深领域主席，International J. Computer Vision、Optimization Methods and Software 编委。



个人简介：王奕森简历：王奕森，北京大学助理教授，博士生导师。研究方向为机器学习理论和算法，重点关注对抗鲁棒性、图学习、弱/自监督学习理论等。目前已发表人工智能领域顶级学术论文 50 余篇，包括 ICML、NeurIPS、ICLR 等，多篇被选为 Oral 或 Spotlight。曾获得 ECML 2021 最佳论文奖。

特邀讲者简介



特邀讲者 1：尤著宏 西北工业大学教授

报告题目：“生物医学知识图谱表示学习：在分子相互作用预测中的应用”

报告摘要：后基因组时代生物医学领域的关键问题之一是系统地理解和分析生物分子之间的分子间关系。复杂的生命活动是由各种生化分子进行的，它们相互作用以维持生物体的正常生理功能，其扰动或破坏将导致功能失调或复杂疾病。在还原论思想的指导下，现有研究对单一或极少数分子的活动进行了深入研究，但对多分子相互作用网络缺乏全局思维，忽视了其整体性和系统性。随着多组学技术、单细胞测序技术等研究方法的进步，大量关于分子间相关性的生物医学大数据迅速积累，这对传统的分析和挖掘方法提出了挑战，也为我们探索复杂的生物分子相关性网络提供了宝贵的数据支持。本报告结合了知识图谱和推荐系统两种人工智能技术，通过学习知识图中各种生物分子的低维表示和与生物分子属性特征的多模态融合，提出了一种基于不同尺度分子关联图的图论预测挖掘方法，本研究将系统地揭示和构建从分子水平到生物医学系统水平的不同尺度的分子关联图。

个人简介：尤著宏，男，博士，西北工业大学计算机学院教授，博士生导师。中国科学技术大学硕士博士，同济大学及香港理工大学博士后，国家“杰出青年”科学基金获得者，中组部第十四批“国家特聘专家”获得者、国家“优秀青年”科学基金获得者及人社部“香江学者”计划获得者。主要从事模式识别、大数据分析 & 生物信息学研究，在 *Transactions on Neural Networks and Learning Systems*、*IEEE Transactions on Knowledge and Data Engineering*、*PLOS Computational Biology*、*Briefings in Bioinformatics* 等国内外重要学术期刊及 *AAAI*、*IJCNN*、*WCCI*、*BIBM* 等国际会议发表研究论文 350 余篇，SCI 索引论文 280 余篇。其中，第一及通信作者 SCI 论文 192 篇，中科院一区论文 62 篇。发表论文被国际著名杂志及会议引用近 16000 次（*Google Scholar*），H 指数 66，单篇最高引用 570 次，单篇引用 10 次以上 220 篇，ESI 高被引论文 10 篇，热点论文 3 篇，连续多次入选爱思唯尔“高被引学者榜单”、斯坦福大学全球前 2% 顶尖科学家等。撰写专著 1 部、申请发明专利 16 项。先后获得了省级自然科学奖一等奖（第 1 完成人）、教育部自然科学奖二等奖（第 2 完成人）、中国自动化学会自然科学二等奖（第 3）等奖励。担任了十多个 SCI 期刊的编委、特约编委和多个国际会议程序委员会成员。



特邀讲者 2：李建欣 北航教授

报告题目：“大数据关联建模及长序预测方法”

报告摘要：大数据蕴含的关联结构、驱动的长时预测，已成为为网络空间安全、重大设施运维等行业领域的共性需求，关键要解决“数据看得懂、行为判得准、演化能预测”等智能计算难题。针对互联网行为数据关联、检测预测的计算需求，构建了图高阶表征-关联异常发现-长序列预测框架，重点研究：如何发现几何特性高阶保持的低失真图表征空间？如何发现高阶复杂关联模式？如何实现长序列高效预测？并基于异构融合体系提升数据计算效率。研发系统已应用于网信、公安领域的异常群体检测等，以及国网、航天的设备运维等业务。

个人简介：李建欣，教授，北航计算学院党委书记，国家杰出青年基金获得者，CMU 机器学习系访问学者，现任 CCF 青工委副主任，大数据专家委常委，CAAI 组织工委副主任、教工委常委等。研究方向为大数据计算、人工智能等，主持了 2030 新一代人工智能重大项目、国家科技支撑课题、NSFC 联合重点基金等，服务国家网信、公安等领域应用，在 TPAMI、TKDE、TC、NeurIPS、KDD 等发表论文百余篇，获 AAAI 2021、IWQoS 2022、DependSys 2017 最佳论文奖，CIKM 2022 最佳论文荣誉提名奖等。曾获国家级科技奖和省部级一等奖等，入选微软学者、北京市科技新星、中国电子学会优秀科技工作者（十佳）、教育部新世纪优秀人才和青年学者，教育部-华为智能基座优秀教师等。



特邀讲者 3: 李霄剑 合肥工业大学副院长

报告题目: “三维感知驱动的微创手术智能化技术研究”

报告摘要: 手术机器人因其具有创伤小、精度高、稳定性强等特点, 成为了当今社会广泛关注的热点。配备于手术机器人上的电子内窥镜受视场角限制, 导致使用者很难获得全面的体内环境信息, 亟待出现更加直观全面的体内环境实时动态三维影像。本报告基于李霄剑研究员在微创手术机器人领域的研究成果, 探讨了三维感知驱动的智能技术应用, 介绍了实时动态深度估计、动态点云柔性配准、手术器械鲁棒检测及自主追踪算法等多项关键技术, 展示了其在临床与动物实验中的应用成果。

个人简介: 李霄剑, 研究员、博士生导师, 合肥工业大学管理学院副院长, 国家自然科学基金优青项目获得者。致力于自动化智能化医疗机器人系统的研究与开发, 研究方向包含体内环境动态三维感知与重构、微创手术机器人自主控制、机器人人机协同操控、微米机器人导航控制等。主持国家自然科学基金重点项目、优青项目、青年项目, 安徽省自然科学基金杰青项目, 安徽省重大专项等, 2019年入选安徽省高层次人才计划。研究成果以第一作者或通信作者身份发表在 *Science Robotics*、*Automatica*、*TRO*、*TII*、*TMech* 等顶级期刊。获省部级科技进步一等奖 2 项。



特邀讲者 4：江劭玮 杭州电子科技大学教授

报告题目：“高通量叠层显微成像技术与应用”

报告摘要：叠层成像作为近年来迅速发展的新型相干衍射成像技术，凭借其无需参考光束、相位定量化、无像差等优势，已成为众多同步辐射 X 射线实验室中不可或缺的成像方式和相干衍射成像领域的热点研究方向。然而，由于成像通量与分辨率较低，叠层成像技术在可见光波段的应用受限。在本报告中，我们将汇报我们在基于可见光波段的高通量叠层显微成像方面取得的一些研究进展，主要包括基于编码照明的频域傅里叶叠层成像技术和基于编码探测的空域编码叠层成像技术。报告将介绍成像的基本原理，并进一步讨论其大视场、高分辨率、高相位灵敏度、低成本等优势特点，以及在生物医学、光学测量等领域的潜在应用。

个人简介：江劭玮，杭州电子科技大学教授，博士生导师，国家级海外高层次青年人才。2014 年获浙江大学学士学位，2021 年获美国康涅狄格大学博士学位。主要研究方向为计算光学成像，包括无透镜成像、超分辨率成像、定量相位成像等。在 *Nature Protocols*、*Nature Reviews Physics* 等国际高水平学术期刊上发表 SCI 检索论文 50 余篇，其中作为第一作者或通讯作者 30 余篇。发表的论文中，7 篇被选为期刊封面论文，7 篇被选为期刊编辑精选论文。相关成果被 *Physics Today* 等 20 余家国内外媒体报导。

专题论坛：可信具身视觉特征学习

(论坛主席：孙仕亮，上海交通大学教授，
赵才荣，同济大学电子与信息工程学院计算机系教授，
朱磊，同济大学研究员，
叶茫，武汉大学教授，
殷俊，上海海事大学信息工程学院副教授)

(10月19日 17:30-19:35，地点：塔里木厅)

时间 10月19日	报告嘉宾姓名、单位、及联系 信息	报告题目
17:30-17:55	吴小俊 江南大学 wu_xiaojun@jiangnan.edu.cn	黎曼流形上的深度学习
17:55-18:20	孟德宇 西安交通大学 dymeng@mail.xjtu.edu.cn	机器学习的数学技术
18:20-18:45	严严 厦门大学 yanyan@xmu.edu.cn	高精度人脸表情识别技术研究
18:45-19:10	邹征夏 北京航空航天大学 zhengxiazou@buaa.edu.cn	生成式遥感基础模型初探
19:10-19:35	徐婧林 北京科技大学 xujinglin@ustb.edu.cn	面向行为理解的细粒度运动分析

论坛介绍

在大模型的推动下，具身智能迎来了新的发展契机。作为实现具身智能的关键技术，具身视觉特征学习在视觉感知与认知过程中，充分融合了智能系统（如机器人、无人车、无人机等）的身体体验、活动与状态，以实现视觉信息的精准获取、深度处理与理解。它着重强调智能本体与环境的动态互动，借助运动与主动感知能力，赋予智能系统理解和诠释视觉世界的新视角。当前，科研人员正致力于构建大规模具身视觉大模型，旨在模拟并预测身体的视觉感知与认知过程。与此同时，他们也在不断寻求更精确地模拟和理解智能系统状态与环境的方法，以及设计更为自然、高效的、更可信的交互方式，让智能系统具有更强的视觉感知和认知能力。为了深入挖掘可信具身视觉特征学习的前沿动态，本次论坛将特邀六位顶尖专家。他们将围绕该领域的最新研究突破、当前技术挑战以及未来发展趋势展开深入交流与探讨，以期推动可信具身视觉特征学习领域的持续发展与突破。论坛组织者包括上海交通大学孙仕亮教授、同济大学赵才荣教授、同济大学朱磊研究员、武汉大学叶茫教授、上海海事大学殷俊副教授。

论坛主席



个人简介：孙仕亮(shiliangsun@gmail.com)，上海交通大学教授，上海市计算机学会人工智能专委会主任，上海市图像图形学学会机器学习专委会主任。本科毕业于北京航空航天大学，博士毕业于清华大学，曾在 University College London, Columbia University 等海外高校从事访问研究。2011 年至 2023 年在华东师范大学任职教授，2018 年至 2023 年担任华东师范大学人工智能研究所常务副所长。主要研究方向为机器学习、人工智能。担任 IEEE TPAMI、IEEE TNNLS、自动化学报等知名期刊编委。迄今发表学术论文 200 多篇，2020 年出版中文教材《模式识别与机器学习》（清华大学出版社）。



个人简介：赵才荣(zhaocairong@tongji.edu.cn)，工学博士。现任同济大学电子与信息工程学院计算机系教授，博士生导师，智能信息处理教研室主任。曾任香港理工大学兼职研究员（2016-2017）。目前担任上海市计算机学会计算机视觉专委会主任，中国图象图形学学会青工委副秘书长，中国自动化学会模式识别与机器智能专业委员会副秘书长，中国人工智能学会粒计算与知识发现专委会常委，中国计算机学会杰出会员，担任 IEEE TMM Guest Editor、IET IP Associate Editor、《中国图象图形学报》、《计算机科学》编委。

主要研究领域：计算机视觉，主要聚焦于智能视频行人分析及其隐私安全研究，重点研究高效可信行人再识别、多模态数据驱动的自动驾驶以及垂直领域模型的知识表示与推理问题。已在 TPAMI、IJCV、IEEE TIP、IEEE TIFS、《中国科学:信息科学》等国内外重要学术期刊及国际学术会议 CVPR, ICML, ICLR, AAAI, ACM MM 等发表论文 50 余篇，受理发明专利 20 余项（授权 14 项），研究成果获 2022 年上海市科技进步一等奖（排名第四），获《中国科学:信息科学》2023 年度热点论文奖。主持国家自然科学基金 4 项，主持国家重点研发计划子课题以及企业横向课题十余项。建设国家级线下精品课程 1 门，上海市精品课程 1 门。指导学生获中国电子学会优秀硕士论文，获上海市计算机学会优秀硕士论文。



个人简介：朱磊(leizhu0608@gmail.com)，同济大学研究员，博士生导师。目前主要研究方向是高效能多模态大模型和自主具身智能。共发表中国计算机学会推荐（下简称 CCF）A 类会议论文、ACM/IEEE 汇刊论文百余篇，主编英文学术专著 2 部，ESI 高被引/热点论文 8 篇，Google 引用 8400 多次，H 因子 47。获得 CCF A 类会议 ACM SIGIR 2019 和 ACM MM 2019 的最佳论文提名，CCF C 类会议 ADMA 2020 最佳论文奖，ChinaMM 2022 最佳学生论文奖，1 篇论文入选 2019 年中国百篇最具影响国际学术论文，拥有授权专利 22 件。担任 ACM TOMM、IEEE TBD、Information Sciences 等多个国际期刊的编委，CCF A 类会议的领域主席（AC）或高级程序委员会委员（SPC），CSIG 青年工作委员会副秘书长。主持/参与基金委青年/面上/重点等 10 余项纵横课题。获得山东省自然科学二等奖（序 1），吴文俊人工智能自然科学二等奖（序 2），ACM 中国 SIGMM 新星奖，中国图象图形学学会（CSIG）青年工作突出贡献奖，连续 4 年入选斯坦福大学发布的全球前 2% 顶尖科学家榜单，获 Reseach.com 发布的 Rising Star of Science Award。指导研究生获中国自动化学会优秀硕士生学位论文奖和入选中国电子学会硕士学位论文激励计划。



个人简介：叶茫(yemang@whu.edu.cn)，武汉大学教授、博士生导师，国家高层次青年人才，中国科协青年托举人才，湖北省高层次人才。主要研究方向人工智能、多媒体检索、多模态理解、联邦学习等，发表国际期刊会议论文 90 余篇，其中第一/通讯作者发表 CCF-A 类论文 50 余篇，谷歌学术引用 8000 余次。担任 CCF-A 类期刊 IEEE TIFS 等期刊编委，多次担任 CVPR、ACM MM、ECCV 等学术会议领域主席。

主持国自科-香港联合基金、湖北省重点研发计划等 10 余项科研项目。获谷歌优秀奖学金、ICCV2021 无人机特定行人检索赛道冠军、2021 年斯坦福排行榜“全球前 2% 顶尖科学家”、2022 年度百度 AI 华人青年学者等荣誉。指导学生获 CCF 优秀大学生、中国电子学会优秀硕士论文和国自科博士科研基金。个人简介：殷俊(junyin@shmtu.edu.cn)，博士，上海海事大学信息工程学院副教授，上海市浦江人才。分别于 2006 年、2011 年获得南京理工大学学士、博士学位，2016~2020 年在华东师范大学从事博士后研究，2019~2020 年在美国匹兹堡大学访问。主要研究方向是多视图学习、对比学习等机器学习理论及应用。目前担任国际期刊 Expert Systems and Applications、Neural Processing Letters 编委、上海市计算机学会人工智能专委会副主任。在 IEEE TKDE、IEEE TNNLS、PR、Information Fusion 等国际期刊以及 AAI、ACM MM 等国际会议上发表论文 50 余篇，获 2023 年度中国自动化学会自然科学一等奖（排名第四）、2023 年度 IEEE TNNLS 杰出论文奖。主持国家自然科学基金面上/青年项目、上海市自然科学基金青年项目、中国博士后科学基金面上项目。

特邀讲者简介



特邀讲者 1: 吴小俊 江南大学教授

报告题目: “黎曼流形上的深度学习”

报告摘要: 由于视觉信息坐落在黎曼流形上, 因此在黎曼流形上进行深度学习算法的研究将是一种本征的深度学习范式。本报告首先对黎曼流形和黎曼流形学习作简单介绍, 然后通过几个案例给出在黎曼流形上设计深度学习算法的思想和方法。本报告还将介绍黎曼深度学习在图像质量增强、目标跟踪与识别、行为识别以及视觉融合等方面的应用研究。

个人简介: 吴小俊(wu_xiaojun@jiangnan.edu.cn), 国际模式识别协会会员(IAPR Fellow)、亚太人工智能协会会员(AAIA Fellow)、国际人工智能产业联盟产业研究院院士(AIIA Fellow)、江南大学至善教授、研究生院院长、Josef Kittler 人工智能研究院院长、教育部装发创新团队负责人、科技部中英人工智能联合实验室主任、教育部/江苏省人工智能国际合作联合实验室主任、2006 年教育部新世纪优秀人才、江苏省 333 工程第一层次人才。从事模式识别与人工智能的研究, 在 TPAMI、IJCV、TIP、CVPR、ICCV、ECCV、IJCAI、AAAI 等期刊和会议上发表学术论文 400 余篇, 出版学术著作 5 本。研究成果获得国内外学术奖励 30 余项, 其中包括多项国际竞赛冠军、国际会议最佳论文奖、IETE Gowri Memorial Award、教育部科技进步一等奖、合作者 Josef Kittler 院士获 2015 江苏省国际科技合作奖和 2016 中国政府友谊奖; 担任了 2 项国家重点研发计划项目首席科学家、3 项国家自然科学基金重点项目和多项 GF 项目负责人。现任 IEEE 智慧城市指导委员会委员、多本国际期刊主编或编委、教育部计算机类教学指导委员会委员、中国图像图形学会理事和江苏省人工智能学会副理事长等职。



特邀讲者 2: 孟德宇 西安交通大学教授

报告题目: “机器学习的数学技术”

报告摘要: 以深度学习/大模型为代表的机器学习方法与技术为当今科技领域的研究焦点。然而,相比技术水平的迅猛提升,机器学习基础理论研究进展远远滞后,大量技术经验发现无法找到理论支撑,以科学性为前提的学科大厦面临危局。重建机器学习理论体系已成为当今重大科技前沿问题。针对这一挑战,本报告将以深度学习三大技术实验现象(任务泛化能力、智能涌现现象、鲁棒-精确悖论)理论内涵作为分析对象,介绍未来机器学习理论可能发展的新型学习、统计与物理理论新框架,以及本课题组分别针对通信、生命、医疗技术领域的场景动态适应、生命智能涌现、智能可靠诊疗三大技术问题的机器学习数学技术方面的初步探索。

个人简介: 孟德宇(dymeng@mail.xjtu.edu.cn), 西安交通大学教授, 博导, 任大数据算法与分析技术国家工程实验室机器学习教研室负责人。发表论文百余篇, 谷歌学术引用超过 29000 次。现任 IEEE Trans. PAMI, NSR 等 7 个国内外期刊编委。目前主要研究聚焦于元学习、概率机器学习、可解释性神经网络等机器学习基础研究问题。



特邀讲者 3: 严严 厦门大学教授

报告题目: “高精度人脸表情识别技术研究”

报告摘要: 人脸表情识别具有重要的研究价值。为了实现有效的人脸表情识别, 学习鲁棒的表情特征至关重要。然而, 人脸表情图像常常受到人脸身份、光照、姿态等因素的影响, 严重干扰了表情特征的提取。同时, 不同表情之间也存在着极大的相似性。上述问题使得人脸表情识别面临极大的挑战, 精度急剧下降。本次报告将介绍我们组最近在人脸表情识别方面的 3 个研究工作。第一个是基于特征解构与重构学习的人脸表情识别方法。第二个是一种基于干扰分离学习的人脸表情识别方法。第三个是一种基于解耦表示学习的人脸表情识别方法。最后, 简要介绍我们在人脸属性识别上的一些研究。

个人简介: 严严(yanyan@xmu.edu.cn), 厦门大学信息学院计算机系教授, 博士生导师, 计算机系教工党支部副书记。现为国际知名期刊 *Neurocomputing* (JCR 2 区)、*Visual Computer* (JCR 2 区) 责任编委 (Associate Editor)、计算机研究与发展 (CCF 中文期刊 A 类) 青年编委、中国图象图形学报 (CCF 中文期刊 B 类) 青年编委、IEEE 高级会员、CCF 高级会员、CCF 计算机视觉专委会委员、福建省计算机学会第九届理事会理事、福建省和厦门市高层次人才 (B 类)。近年来在 *IEEE T-PAMI* (JCR 1 区, CCF A 类)、*IJCV* (JCR 1 区, CCF A 类)、*IEEE T-Cyber*、*IEEE T-IP*、*IEEE-TMM*、*IEEE T-AC*、*IEEE T-ITS*、*PR*、*中国科学-信息科学*、*计算机学报*、*自动化学报* 等国内外著名期刊以及 *ICCV* (CCF A 类)、*CVPR* (CCF A 类)、*ACM MM* (CCF A 类)、*AAAI* (CCF A 类)、*ECCV* (CCF B 类) 等国际著名会议上发表论文 100 余篇 (JCR 3 区以上 SCI 论文 50 余篇, CCF A 类和 *IEEE Trans* 论文 40 余篇), 其中以第一作者或通信作者发表 SCI 论文 50 余篇 (其中 JCR 2 区以上论文 30 篇以上), 参与出版人脸识别学术专著一部。此外, 已获批准发明专利 30 余项。



特邀讲者 4：邹征夏 北京航空航天大学教授

报告题目：“生成式遥感基础模型初探”

报告摘要：遥感基础模型（大模型）作为遥感技术与智能前沿的交叉融合，是空天信息领域的基础前沿方向。生成式遥感基础模型作为其中的一类重要分支，在数字化遥感场景构建、赋能下游遥感图像解译等任务都起着关键的作用。在本次报告中，讲者将介绍课题组近期在遥感生成式基础模型方面的工作，包括全球范围遥感图像生成、多模态遥感图文生成。最后，对生成式遥感基础模型的潜在应用场景和未来研究方向进行介绍。

个人简介：邹征夏，北京航空航天大学宇航学院教授、博士生导师，国家级青年人才。主要研究方向包括遥感图像处理、计算机视觉、深度学习等，研究成果以第一/通讯作者身份发表在 Proceedings of the IEEE、Nature Communications（首页论文）、IEEE Transactions 汇刊、CVPR、ICCV 等重要期刊和会议，谷歌学术论文引用 6000 余次，单篇论文引用 2000 余次。入选 2022/2023 全球前 2% 科学家名单，担任 Nature 旗下期刊 Communications Engineering 特刊编辑，研究成果收录于斯坦福大学著名公开课，被新华社、中央电视台、新科学人等媒体报道和采用，服务于航天国防、自动驾驶等重要应用。



特邀讲者 5：徐婧林 北京科技大学副教授

报告题目：“面向行为理解的细粒度运动分析”

报告摘要：细粒度运动分析旨在通过精细化地分析人体动作序列，实现动作的细粒度识别、定位、姿态估计、质量评价，可广泛应用于智能安防、智慧医疗、智能体育、智慧传媒等领域。本报告首先介绍细粒度运动分析面临的挑战与意义，重点介绍文本驱动的细粒度时空动作定位、细粒度提示驱动的三维人体姿态估计、以人为中心的细粒度动作质量评价，回答如何在时间和空间上定位边界模糊的细粒度动作、如何在三维空间中估计深度不确定的人体姿态、如何从细粒度层面更精确地评价人体动作质量？上述研究工作将在运动康复、体能测试、数字传媒等领域具有重要意义。

个人简介：徐婧林(xujinglin@ustb.edu.cn)，北京科技大学智能科学与技术学院副教授，北京图象图形学学会理事、副秘书长。入选第九届中国科协青年人才托举工程、2022年获中国图象图形学学会优秀博士学位论文奖、2023年获中国自动化学会自然科学奖一等奖（4/5）。主要研究方向为视频动作理解、多模态细粒度分析、三维人体姿态估计与动作生成。已发表/接收 18 篇 CCF A 类国际期刊和会议论文（一作 14 篇，通讯 1 篇）。代表性一作论文包括：CVPR 论文 5 篇（2 篇 Oral/1 篇 Spotlight/1 篇 Highlight）、IEEE TIP 论文 3 篇、IEEE TKDE 论文 1 篇、AAAI 和 IJCAI 论文 5 篇。主持国家自然科学基金面上项目、青年基金、中国博士后科学基金面上项目等。获得北京图象图形学学会“最美女科技工作者”称号、2022 年西北工业大学优秀博士学位论文、2020 年西北工业大学优秀毕业生等荣誉。担任《电子与信息学报》编委、《计算机科学》青年编委、《人工智能》编委等。

专题论坛：视言碰撞：语言模型与视觉生态协同

（论坛主席：程明明，南开大学教授，

南开大学，博士生导师，

贾旭，大连理工大学副教授）

（10月19日 17:30-19:30，地点：报告厅）

时间	报告嘉宾姓名、单位、及联系信息	报告题目
17:30-17:50	程明明 南开大学 cmm@nankai.edu.cn	开场白
17:50-18:10	徐畅 悉尼大学 c.xu@sydney.edu.au	Pixels as Parameters in Diffusion Models
18:10-18:30	任博 南开大学 rb@nankai.edu.cn	含复杂光路场景的三维神经渲染
18:30-18:50	盛律 北京航空航天大学 lsheng@buaa.edu.cn	Multimodal LLMs Meet Embodied Agents
18:50-19:10	王毅 上海人工智能实验室 wangyi@pjlab.org.cn	InternVideo2: Scaling Video Foundation Models for Multimodal Video Understanding
19:10-19:30	杨灵 北京大学 yangling@stu.pku.edu.cn	扩散模型的前沿算法与应用

论坛介绍

随着语言与视觉大模型的迅猛发展，语言模型与视觉生态紧密地交互，诞生出了空前的探索机遇，也引领了当下研究的前沿与潮流。本论坛从语言模态和视觉模态的碰撞出发，重点探讨语言模型对于视觉生成、感知、理解乃至交互多个层面的协同与促进作用，希望对这一领域的研究和发展做出积极的贡献。具体地，本次论坛特邀讲者的报告内容主要涵盖（1）如何修正语言模型获取高质量的图像生成及 2D/3D 图像生成模型；（2）如何利用语言模型进行视觉任务的调度；（3）如何掌控 3D 视觉交互与控制。我们期待在本论坛中与各位研究者和从业者共同探讨，推动这一领域的发展，创造出更多创新的应用，激发出更多的创造潜力。论坛预计 200 人。

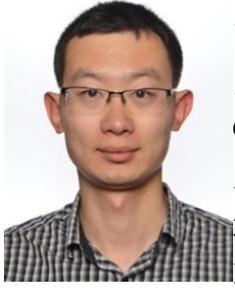
论坛主席



个人介绍：程明明(cmm@nankai.edu.cn)，南开大学杰出教授，主持承担了国家杰出青年科学基金、优秀青年科学基金项目、科技部重大项目课题等。他的主要研究方向是计算机视觉和计算机图形学，在 SCI 一区/CCF A 类刊物上发表学术论文 100 余篇（含 IEEE TPAMI 论文 30 余篇），h-index 为 80，论文谷歌引用 4 万余次，单篇最高引用 4700 余次，多次入选全球高被引科学家和中国高被引学者。技术成果被应用于华为、国家减灾中心等多个单位的旗舰产品。获得教育部自然科学一等奖 2 项、其他省部级科技奖 2 项。培养的 3 名博士生获得省部级优秀博士论文奖。现担任中国图象图形学学会副秘书长、天津市人工智能学会副理事长和顶级期刊 IEEE TPAMI, IEEE TIP 和《中国科学：信息科学》编委。



个人介绍：王亚星(yaxing@nankai.edu.cn)，南开大学，博士生导师，入选海外高层次人才引进计划青年项目，入选南开“百名青年学科带头人培养计划”。西班牙巴塞罗那自治大学博士，曾在西班牙巴塞罗那自治大学从事博士后研究。研究方向为扩散模型、生成对抗网络、图像到图像翻译、迁移学习。在 TPAMI, IJCV, CVPR, NeurIPS 等期刊会议发表论文 30 余篇，谷歌学术引用 2000 余次。现担任 Computers, Materials & Continua 期刊编委，ECCV Workshop 组织者，在国际顶级期刊和会议 TPAMI、NeurIPS、CVPR、ICCV 等多次担任期刊和会议审稿人。多模态语言翻译国际竞赛 (WMT16 Multimodal Machine Translation challenge) 中 荣获第一名、2022 年粤港澳大湾区（黄埔）国际算法算例大赛（遥感目标检测赛道）亚军（2/116 队伍）。主持国家自然科学基金青年项目。



个人介绍：贾旭(xjia@dlut.edu.cn),大连理工大学未来技术学院/人工智能学院院长聘副教授,辽宁省智能感知与理解人工智能重点实验室骨干成员,博士毕业于比利时鲁汶大学,师从 Tinne Tuytelaars 教授和 Luc Van Gool 教授,曾在 Google Research, 商汤科技, 华为诺亚方舟实验室等从事研究工作。现主要研究方向包括视觉内容增强与生成、类脑视觉等,近年来在计算机视觉和机器学习领域顶级会议及期刊发表论文 40 余篇, Google Scholar 引用 8300 余次,申请国内外专利 10 余项。主持或参与国家自然科学基金重点项目、科技部科技创新 2030 重大项目以及华为等多项科研项目。担任 IJCAI、ICLR 多个国际顶级会议和期刊的领域主席和审稿人, CCF、CSIG 中多个专委会执委, 及 VALSE 第六、七届执委。

特邀讲者简介



特邀讲者 1: 徐畅 悉尼大学副教授

报告题目: Pixels as Parameters in Diffusion Models

报告摘要: 在当今 AI 生成内容的世界中，扩散模型作为一种有价值的技术正越来越受到关注。在本次报告中，我们将介绍扩散模型及其在计算机视觉中的相关应用。这些模型不仅限于生成图像，还可以应用于各种计算机视觉任务。例如，我们将讨论我们最近在视频动作分割中使用扩散模型的工作。此外，可以将反向扩散过程视为逐步更新图像像素的一种方式，实际上将像素视为参数或变量。这一视角为优化扩散模型中的采样过程开辟了一种全新的方法。通过重新思考如何将像素更新为参数，我们发现了一些可能阻碍图像生成稳定性的有趣挑战，并为此开发了创新的优化策略来应对这些挑战。

个人介绍: 徐畅，澳大利亚悉尼大学副教授，澳大利亚研究理事会杰出青年学者（ARC Future Fellow）。他的研究领域主要集中在机器学习算法及相关计算机视觉应用。他已在国际知名期刊和顶级学术会议上发表了超过 200 篇论文，并获得多项重要的学术奖项，包括 2023 年 AAAI 杰出论文奖，以及 2018 年 IJCAI 杰出论文奖。他在 NeurIPS、ICML、ICLR 和 CVPR 等知名会议担任领域主席，并在 AAAI 与 IJCAI 担任高级程序委员。此外，他还担任 IEEE T-PAMI、IEEE T-MM 和 T-MLR 的副主编。他荣获 2023 年度澳大利亚新南威尔士州州长优秀青年研究奖。



特邀讲者 2：任博 南开大学副教授

报告题目：含复杂光路场景的三维神经渲染

报告摘要：高质量的大模型训练需要高质量的数据，这在三维方面尤其如此。然而，当前的三维数据质与量远达不到要求。使用计算手段高效生成真实高质量的三维数据是一种可行的方案。神经渲染技术在利用图像与视频生成三维数据方面在近年间展示了较强的潜力，然而其在包含多镜面、半透明、运动烟雾等导致复杂光路的场景尚存在较大的挑战。多数相关的研究局限于处理不透明的二维流形。在本次报告中，我将探讨如何从第一性原理和基础假设出发增强神经渲染理论，使之能够胜任更广泛的场景数据生成任务。

个人介绍：任博，南开大学计算机学院副教授。主要研究方向包括计算机图形学基于物理/机器学习的仿真与控制，神经辐射场三维场景重建与渲染等。在国际顶级期刊会议发表文章二十余篇。主持或参与多项国家自然科学基金青年/面上项目，国家重点研发计划课题。任中国图学学会理事会国际联络工作委员会，CCF CAD&CG 专委会，CSIG 智能图形专委会委员。在 SIGGRAPH Asia, CVM, Pacific Graphics 等图形学国际会议中出任分会场主席。



特邀讲者 3: 盛律 北京航空航天大学副教授

报告题目: Multimodal LLMs Meet Embodied Agents

报告摘要: 在多模态大模型和生成模型的加持下,围绕具身智能体的研究工作从传统的抓取、导航等任务,扩展到在复杂多模态环境中通用交互任务的结果。在本次报告中,讲者将介绍面向具身智能体的多模态大模型及其多模态多任务高效微调方法,并以此为基础探讨实现具身智能体泛化能力的几种设计思路,包含 Minecraft 等仿真环境和机器臂操作等真实环境下的尝试。最后,报告将展望端到端具身大模型,讨论其中的机会和挑战。

个人介绍: 盛律(lsheng@buaa.edu.cn), 博导,北京航空航天大学“卓越百人”副教授,入选北航青年拔尖计划。主要研究方向为三维视觉和具身智能。在 IEEE TPAMI/IJCV/TIP 以及 CVPR/ICCV/NeurIPS/ICLR/ECCV 等重要国际期刊和会议发表论文超过 50 篇,Google Scholar 显示被引用数超 5000 次。组织 ICML 2024 Multimodal Foundation Models Meet Embodied AI 和 ICCV 2021 SenseHuman 等多个国际会议研讨会。现任 ACM Computing Surveys 副编辑, CVPR 2024、ECCV 2024 和 ACM Multimedia 2024 领域主席,以及多个领域顶会顶刊审稿人和程序委员。任 CCF 和 CSIG 多个专委会执行委员, VALSE 执行领域主席。主持或参与多项国家自然科学基金、科技部重点研发计划和省部级重点研发计划项目。



特邀讲者 4: 王毅 上海人工智能实验室青年研究员

报告题目: InternVideo2: Scaling Video Foundation Models for Multimodal Video Understanding

报告摘要: 视频理解是计算机视觉领域的关键研究方向,拥有广泛的应用场景,例如视频搜索、游戏控制、机器人学习、自动驾驶和科学研究等。近年来,大型语言模型 (LLM) 和多模态大型语言模型 (MLLM) 的发展对视觉研究和其他学科产生了深远的影响。将视频有效地嵌入这些大型模型中,并利用其能力来增强视频理解性能,已成为一项关键任务。我们引入 InternVideo2, 一个新的视频基础模型 (ViFM) 家族,它在视频识别、视频文本任务和以视频为中心的对话中取得了最先进的结果。我们的核心设计是一种渐进式训练方法,它将掩码视频建模、跨模态对比学习和下一个令牌预测统一起来,将视频编码器的规模扩展到 60 亿个参数。在数据层面,我们通过语义分割视频和生成视频音频语音字幕来优先考虑时空一致性。这提高了视频和文本之间的对齐。通过大量的实验,我们验证了我们的设计,并在 60 多个视频和音频任务上展示了优越的性能。值得注意的是,我们的模型在各种与视频相关的对话和长视频理解基准测试中优于其他模型,突出了其推理和理解更长上下文的能力。

个人介绍: 王毅(wangyi@pjlab.org.cn),在香港中文大学获得博士学位,专注于计算机视觉中的视频/图像理解和生成。他在顶级期刊和会议等发表 20 余篇论文,1 篇论文曾入选 CVPR2022 best paper finalist。他在多个期刊和会议上担任评审。他的研究工作获得了超过 2500 次引用。他曾获得 9 项国际比赛第一名,包括 CVPR 具身智能 RxR-Habitat 赛道冠军,ECCV 第一视角视频 5 个赛道冠军等。成果包括通用视频大模型 InternVideo、视频多模态数据集 InternVid 和首个视频对话系统 VideoChat。



特邀讲者 5: 杨灵 北京大学博士在读

报告题目: 扩散模型的前沿算法与应用

报告摘要: 扩散模型 (Diffusion Models) 作为生成模型的一种新兴方法, 在图像生成领域取得了显著的成果。扩散模型通过逐步添加噪声并学习逆过程生成数据, 展现了其在高质量图像生成中的潜力。然而, 如何在生成过程中实现精确的可控性, 仍然是一个具有挑战性的问题。本报告将探讨扩散模型的可控生成、复杂生成机制, 以及其在图像、视频、3D 和 4D 等多领域的应用。首先, 我将介绍扩散模型的一些可控生成方法, 通过引入条件信息或引导策略, 实现对生成结果的精确控制。随后, 我将展示基于 (多模态) 大语言模型的复杂生成框架和优化策略, 使扩散模型在生成过程中能够处理更复杂的场景。最后, 我将展示扩散模型在图像、视频、3D 和 4D 等领域的应用方法和实例。

个人简介: 杨灵, 北京大学博士在读, 导师为崔斌教授, 研究方向为扩散模型, 多次获得北京大学国家奖学金、学术创新奖等, 入选 Valse 2024 优秀学生论坛 (全国一共 8 名学生)。主编 AIGC 专著《扩散模型: 生成式 AI 模型的理论、应用与代码实践》, 一作在 CVPR/NeurIPS/ICML/ICLR/TKDE 等顶刊顶会上共发表论文 15 篇, 长期担任 SIGGRAPH, TPAMI, ICML, ICLR, NeurIPS, CVPR 等顶刊顶会审稿人, 一作发表文生图 SOTA 框架 RPG-DiffusionMaster, 性能超越 Stable Diffusion XL 和 OpenAI 的 DALL-E 3, 和 OpenAI 合作一作发表全球首篇扩散模型综述。长期和 OpenAI、斯坦福大学等知名研究机构在文生图/视频扩散模型等研究领域进行合作探索, 一作文章总被引 1300 次。

专题论坛：开放环境可信感知模型和学习

(论坛主席：刘成林，中国科学院自动化研究所副所长，

胡清华，天津大学北洋讲席教授，

朱飞，中国科学院香港创新研究院人工智能与机器人创新中心博士后)

(10月19日 17:30-19:40，地点：学术厅)

时间	报告嘉宾姓名、单位、及联系信息	报告题目
10月19日		
17:30-18:00	龚怡宏 西安交通大学 ygong@mail.xjtu.edu.cn	受脑认知启发的小样本持续学习方法
18:00-18:30	朱军 清华大学 dcszj@mail.tsinghua.edu.cn	从生成模型的角度探索高效鲁棒的模式识别方法
18:30-19:00	程明明 南开大学 cmm@nankai.edu.cn	开放动态场景下新类别发现与更新
19:00-19:20	张长青 天津大学 zhangchangqing@tju.edu.cn	低质多模态数据融合
19:20-19:40	张煦尧 中科院自动化所 xyz@nlpr.ia.ac.cn	模式识别置信度估计

论坛介绍

随着人工智能技术的发展，深度模型在各种感知与学习任务上取得了显著的性能突破。然而，这些优异的表现大多依赖于静态、封闭的测试环境。在真实的开放环境中，面对动态、复杂的输入数据，深度模型存在诸多不足：一方面，模型往往对其预测过度自信，既不能对已知类别给出可靠的置信度，也不能拒识未知类别或噪声模式，给实际应用带来了安全隐患；另一方面，在新任务上更新模型往往导致其对旧知识的灾难性遗忘，缺乏持续学习的能力。近年来，越来越多的研究人员开始关注开放环境下模型的可靠性和动态自适应能力，并提出了一系列创新的方法。预训练模型和多模态大模型的发展给可信学习和持续学习带来了新的机遇和挑战。本论坛旨在汇聚开放环境可信感知与学习领域的专家学者，深入探讨开放环境下模型置信度估计、持续学习、多模态基础模型可信学习等方面的最新进展、应用实践和发展趋势。

论坛主席



个人简介：刘成林(liucl@nlpr.ia.ac.cn)，中国科学院自动化研究所副所长，研究员、博士生导师，中国科学院大学人工智能学院副院长。1989年、1992年、1995年分别在武汉大学、北京工业大学、中国科学院自动化研究所获学士、硕士和博士学位。1996年至2004年先后在韩国科学技术院、日本东京农工大学、日立中央研究所从事博士后和研发工作。2005年起在中国科学院自动化研究所任研究员。2008年获得国家杰出青年科学基金。研究兴趣包括模式识别、机器学习、文字识别与文档分析等。在国内外期刊和学术会议上发表论文400余篇，合著英文专著一本。现任 *Pattern Recognition* 期刊和《自动化学报》的副主编，以及多个期刊的编委。任中国自动化学会会士、模式识别与智能系统专委会主任，中国人工智能学会副理事长、会士，美国电气电子工程师协会会士（IEEE Fellow）、国际模式识别学会会士（IAPR Fellow）。



个人简介：胡清华(huqinghua@tju.edu.cn)，国家优青/杰青，天津大学北洋讲席教授，天津市机器学习重点实验室主任，城市智能与数字治理教育部工程研究中心主任，CAAI 粒计算与知识发现专委会主任。从事大数据粒计算、多模态学习、不确定性建模和自主机器学习方面的研究，先后获得国家重点研发计划项目、国家自然科学基金重点项目、国家优青/杰青以及国防项目的资助。在 *IEEE-TPAMI*、*IJCV*、*IEEE-TKDE*、*IEEE-TFS* 等期刊以及 *NeurIPS*、*CVPR*、*IJCAI*、*AAAI* 等会议发表论文300余篇，获批准发明专利30余个，出版专著3部，先后获得黑龙江省自然科学一等奖和天津市科技进步一等奖。目前担任 *IEEE Trans. Fuzzy Systems*，*自动化学报*、*电子学报*、*智能系统学报* 等期刊的编委。



个人简介: 朱飞(zhufei2018@ia.ac.cn), 中国科学院香港创新研究院人工智能与机器人创新中心博士后。2018 年和 2023 年分别在清华大学、中国科学院自动化研究所获学士和博士学位。曾获中国科学院院长特别奖和中国科学院优秀博士学位论文奖。研究兴趣包括模式识别、机器学习、持续学习与不确定性估计等, 在 IEEE-TPAMI、IJCV、Neural Networks、Pattern Recognition、自动化学报等期刊与 CVPR、ECCV、NeurIPS 等会议上发表论文多篇。

特邀讲者简介



特邀讲者 1：龚怡宏 西安交通大学软件学院院长

报告题目：“受脑认知启发的小样本持续学习方法”

报告摘要：当前的深度学习神经网络（DNN）在增量学习新知识时发生“灾难性遗忘”旧知识的问题，不具备连续学习的能力。基于 ViT 的图像识别大模型由于其网络结构更加复杂，在学习新知识时会产生更加严重的灾难性遗忘现象。本团队通过大量实验研究进一步发现：主流的基于知识蒸馏的持续学习方法存在“旧知识遗忘后重学习”问题，降低了持续学习的精度与收敛速度。本团队与认知科学领域的专家学者开展长期深层次的合作，通过借鉴人脑视觉记忆与概念学习的工作机理，取得以下两个代表性研究成果：
（1）提出基于人脑视觉工作记忆拓扑不变性的持续学习方法，推导出该方法的稳定性与可塑性数学分析结果；
（2）提出基于人脑类比学习机制的小样本新概念学习与泛化方法。大量对比实验展示上述两个方法解决灾难性遗忘问题的有效性和卓越的持续学习性能。

个人简介：龚怡宏(ygong@mail.xjtu.edu.cn)，西安交通大学软件学院院长，西安交通大学软件学院院长，国家特聘教授，IEEE Fellow，国家 973 项目首席科学家，视觉信息处理国家工程实验室副主任，陕西省人工智能联合实验室执行副主任。共出版英文专著 2 部，中文编著 1 部，在国际知名学术期刊和会议上发表论文 200 余篇，谷歌学术引用 3 万余次，谷歌 h-index 为 78。于 ACM SIGIR 2003 会议上发表的文章获得最有影响力文章荣誉提名奖。拥有美国和中国发明专利 30 余项。研究领域包括人工智能，计算机视觉、多媒体内容分析、受脑启发神经网络等，是学术界最早开展体育视频内容分析，图像内容检索，以及提出图像稀疏编码特征向量的国际知名学者，引领了国际上大量学者的研究，带领团队多次获得国内外图像视频内容分析领域技术大赛的冠军。



特邀讲者 2: 朱军 清华大学教授

报告题目: “从生成模型的角度探索高效鲁棒的模式识别方法”

报告摘要: 判别式与生成式是模式识别的两种基本范式, 传统的生成模型受限于模型表达能力, 通常引用过高的偏差。近期, 得益于深度生成模型的快速进展, 需要重新审视生成式模式识别的前景。该报告将介绍深度生成模型(如扩散模型)的前沿进展, 以及基于深度生成模型的高效鲁棒的模式识别方法。

个人简介: 朱军(dcszj@mail.tsinghua.edu.cn), 清华大学计算机系博世人工智能冠名教授、IEEE/AAAI Fellow、清华大学人工智能研究院副院长、计算机系人智实验室主任, 曾任卡内基梅隆大学兼职教授。2001-2009 年获清华大学学士和博士学位, 主要从事机器学习研究, 发表 CCF A 类会议/期刊论文百余篇, 谷歌学术引用 3 万余次; 担任国际著名期刊 IEEE TPAMI 的副主编, 担任 ICML、NeurIPS、ICLR 等(资深)领域主席 20 余次; 曾获中国科协求是杰出青年奖、科学探索奖、中国计算机学会自然科学一等奖、吴文俊人工智能自然科学一等奖、ICLR 国际会议杰出论文奖等, 入选国家级高层次人才计划、中国计算机学会青年科学家、MIT TR35 中国先锋者、IEEE AI 10 to Watch 等。



特邀讲者 3: 程明明 南开大学教授

报告题目: “开放动态场景下新类别发现与更新”

报告摘要: 在开放动态场景中, 新类别的发现与更新是当前机器学习和人工智能领域的一项重要挑战。本报告旨在探讨如何在复杂和变化的环境中, 通过利用大量合成视觉类别来学习未来新类别的多样化和判别性表示, 结合基于最近邻距离感知标签一致性样本选择的渐进监督表示学习, 以增强新类别的发现能力。此外, 目前的新类别发现方法大多仅使用单一的视觉模态信息, 我们进一步探索在图像文本预训练模型中引入文本信息, 以提供补充的判别信息。最后, 我们对新类别发现与更新的研究进行总结和展望。

个人简介: 程明明(cmm@nankai.edu.cn), 南开大学杰出教授, 新一代人工智能发展战略研究院副院长。主持承担了国家杰出青年科学基金、优秀青年科学基金项目、科技部重大项目课题等。他的主要研究方向是人工智能、计算机视觉和计算机图形学, 在 SCI 一区/CCF A 类刊物上发表学术论文 100 余篇 (含 IEEE TPAMI 论文 38 篇), h-index 为 93, 论文谷歌引用 5.4 万余次, 单篇最高引用 5 千余次, 多次入选全球高被引科学家和中国高被引学者。技术成果被应用于华为、国家减灾中心等多个单位的旗舰产品。获得教育部自然科学一等奖 2 项、其他省部级科技奖 2 项。培养的 3 名博士生获得省部级优秀博士论文奖。现担任中国图象图形学学会副秘书长、天津市人工智能学会副理事长和顶级期刊 IEEE TPAMI, IEEE TIP 和《中国科学: 信息科学》编委。



特邀讲者 4: 张长青 天津大学教授

报告题目: “低质多模态数据融合”

报告摘要: 多模态信息融合在科学发现、医疗诊断、机器人等领域具有广泛和重要应用。不同模态数据蕴含的信息具有互补性、冗余性、动态性、不平衡、不完整等诸多复杂关联和不确定性, 其对多模态数据融合的效果产生了深刻的影响。本报告将从统一的视角对低质量多模态数据进行刻画, 并介绍面向低质量多模态数据的融合理论、方法和应用。

个人简介: 张长青(zhangchangqing@tju.edu.cn), 天津大学智能与计算学部教授/博士生导师、人工智能学院副院长, 国家“万人计划”青年拔尖人才, 其主要研究方向为机器学习与计算机视觉。在 Nature Communications/TPAMI/ICML 等顶级期刊和国际会议上发表论文 50 余篇, 多篇论文入选 ICML/CVPR/NeurIPS 口头报告或亮点论文, 谷歌学术引用 1 万余次。研究成果获得重庆市自然科学一等奖、中国图象图形学学会自然科学奖一等奖、ICME 最佳论文等奖励, 入选百度发布的全球高潜力 AI 华人青年学者榜单、斯坦福大学发布的全球 Top 2% 顶尖科学家榜单、爱思唯尔“中国高被引学者”。受邀担任 Pattern Recognition 副编辑及顶级会议 ICLR 领域主席。



特邀讲者 5: 张煦尧 中国科学院研究员

报告题目: “模式识别置信度估计”

报告摘要: 在模式识别问题中除了获得较高的分类精度外, 如何获取有效的置信度估计是另一个重要的问题。置信度反应了识别结果的可靠性, 通常可用于错分样本检测以及分布外样本检测等任务, 对于提升开放环境下的鲁棒性具有重要意义。该报告将介绍课题组在置信度估计方面的最新科研进展, 并对未来研究趋势进行展望。

个人简介: 张煦尧(xyz@nlpr.ia.ac.cn), 中国科学院自动化研究所研究员, 国家优青, 科技创新 2030 新一代人工智能重大项目青年科学家。2013 年获中科院自动化所模式识别与智能系统博士学位, 中国计算机学会优秀博士论文。2012 年加拿大模式识别与机器智能中心访问学者 (Ching Y. Suen)。2015 年深度学习发源地蒙特利尔大学访问学者 (Yoshua Bengio)。主要研究兴趣包括模式识别与机器学习, 在顶级期刊和会议上发表论文 80 余篇, 包括 Proceedings of the IEEE, IEEE TPAMI, CVPR, ICCV 等。担任国际期刊 IEEE TIP 和 Pattern Recognition 的编委。入选中国科学院稳定支持基础研究领域青年团队、中国科协青年人才托举工程、中国科学院青年创新促进会。荣获吴文俊人工智能优秀青年奖、中国科学院朱李月华优秀教师奖。

专题论坛：遥感图像智能解译与应用

(论坛主席：贺威，武汉大学测绘遥感信息工程国家重点实验室教授，

董燕妮，武汉大学资源与环境科学学院教授，

张磊，西北工业大学计算机信息工程系主任，

高常鑫，华中科技大学教授，

程焱，西北工业大学教授，

李淼，国防科技大学电子科学学院特聘副研究员)

(10月20日 17:30-19:30，地点：多功能厅)

时间	报告嘉宾姓名、单位、及联系信息	报告题目
10月20日 17:30-18:00	吴立刚 哈尔滨工业大学 ligangwu@hit.edu.cn	空间非合作目标的智能感知与控制
18:00-18:30	高连如 中国科学院空天信息创新研究院 gaolr@radi.ac.cn	高光谱遥感图像目标探测技术
18:30-19:00	孙显 中国科学院空天信息创新研究院 sunxian@aircas.ac.cn	分布式遥感智能基础模型技术进展及实践
18:30-19:00	李盛阳 中国科学院大学 shyli@csu.ac.cn	空间对地观测视频智能处理-挑战、数据集与方法
19:00-19:30	杨文 武汉大学 yangwen@whu.edu.cn	高分光学遥感图像小目标检测

论坛介绍

遥感影像智能解译是人工智能、计算机视觉等理论应用于遥感观测的核心关键技术。以深度学习为代表的人工智能技术手段为遥感影像智能解译提供了高效、准确、稳定的解决方案，推进了大规模遥感对地观测应用以及深空探测的发展。一方面，遥感智能解译与应用已是计算机视觉领域中不可或缺的研究热点。另一方面，遥感观测与联合国确定的 40 项可持续发展目标中的 36 项有直接联系，是监测、解决人类社会面临的土地、资源、深空环境等关键问题的重要手段。本专题论坛将围绕遥感信息智能提取、遥感微小目标识别、深空探测成像等关键主题，邀请的专家、学者分享并讨论“AI+遥感”的最新研究与未来发展。

论坛主席



个人简介：贺威，武汉大学测绘遥感信息工程国家重点实验室教授、博士生导师，国家优秀青年基金（海外）获得者、湖北百人。主要从事人工智能、遥感图像处理与应用的研究。已发表 IEEE TPAMI、TIP、TGRS、ISPRS 等 SCI 期刊论文 50 余篇，ESI 热点论文 2 篇（地学前 0.1%）、ESI 高被引论文 11 篇（地学前 1%）；CVPR 等会议论文 7 篇，Google Scholar 总引用 4700 余次。研究成果获日本 RIKEN 研究激励奖、日本遥感协会最佳论文奖、IEEE 国际地球学科与遥感协会最佳论文奖和 IEEE 国际地球科学与遥感协会数据融合竞赛双赛道冠军等。主持日本 JSPS 若手研究项目 1 项、自然科学基金面上项目 1 项。获 IEEE TGRS 期刊最佳审稿人称号，测绘学报、Geo-spatial Information Science 期刊青年编委。



个人简介：董燕妮，武汉大学资源与环境科学学院教授、博士生导师，国家优秀青年科学基金获得者。主要从事遥感图像智能解译与地学应用等方面的研究，已在 IEEE TIP、TCybern 和 TGRS 等国际刊物上发表 SCI 检索论文四十余篇。主持国家自然科学基金优青项目、面上项目、青年项目等十余项科研项目。目前担任 SCI 期刊 IEEE Geoscience and Remote Sensing Letter 副主编、Remote Sensing 客座编辑等；曾获得中国图象图形学学会石青云女科学家奖、香江学者奖、中国图象图形学学会优秀博士学位论文提名奖等。



个人简介：张磊，教授博导，国家级青年人才、中国科协“青托”，西北工业大学计算机信息工程系主任。长期围绕太空安全、生态保护等国家重大战略问题，深入研究遥感光学计算成像与图像处理技术。发表论文 100 余篇，包括中科院 SCI 分区一区 13 篇，中国计算机学会 (CCF) 推荐 A 类 20 篇，IEEE 会刊 20 篇。获国家发明专利 24 项，参与制定团体标准 1 项。主持国家级项目 5 项，省部级项目 2 项。获中国图象图形学学会 (CSIG) 优秀博士论文奖、ICOT2019 最佳论文奖。担任 CSIG 成像探测与感知专委会秘书长、CAA 模式识别与机器智能专委会委员。



个人简介：高常鑫，华中科技大学教授、多谱信息智能处理技术国家重点实验室副主任，兼任中国图象图形学学会成像探测与感知专委会副秘书长、中国人工智能学会模式识别专委会副秘书长。研究领域为计算机视觉、图像/视频智能理解、语义分割、行为识别，及其在视频监控等领域的应用。近年来发表 IEEE TPAMI、IJCV、IEEE TIP、CVPR、ICCV、ECCV 等高水平期刊会议论文 40 余篇。先后主持湖北省杰出青年基金项目、国家自然科学基金面上项目、装备预研项目、科技部重点研发子课题等项目。获湖北省科技进步一等奖。



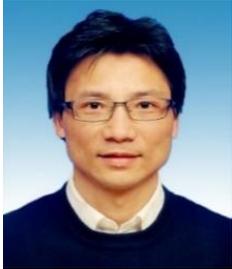
个人简介：程焱，西北工业大学教授，博士生导师，信息融合技术教育部重点实验室副主任，国家级青年人才，连续 4 年入选科睿唯安“全球高被引科学家”和爱思唯尔“中国高被引学者”。主要研究方向为光学遥感图像理解、计算机视觉等。以第一作者/通讯作者发表论文 60 余篇，包括 Proceedings of the IEEE、TPAMI、CVPR、ICCV 等，30 余篇论文入选 ESI 高被引论文/热点论文，谷歌学术总引用 2 万余次，5 篇论文单篇引用大于 1000 次，获得 2021 年度 IEEE TCSVT 最佳论文奖、2021 年度和 2023 年度 IEEE 地球科学与遥感学会最高影响力论文奖等学术奖励，获得中国图象图形学学会青年科学家奖、中国电子学会自然科学一等奖、陕西省科学技术一等奖等科技奖励，担任 IEEE GRSM、IEEE TGRS、ISPRS JPRS、JRS 等多个国际期刊编委。



个人简介：李森，国防科技大学电子科学学院，特聘副研究员，主要研究远距离下的空天时敏目标智能检测、动态事件感知一体化机制、基于类脑计算的轻量化检测等，作为目标检测技术负责人，致力于最新研究成果与国家重大需求紧密结合，解决多个地面和星上信息处理系统的极弱、超快、低虚警检测难题，获军队科技进步一等奖 1 项、军队科技进步二等奖 1 项，主持和参与国家自然科学基金、装备预研、演示验证、自主科研项目等 12 项，发表 TPAMI、TIP 等高水平学术

论文 20 篇，申请专利 10 项。

特邀讲者简介



特邀讲者 1：吴立刚 哈尔滨工业大学教授

报告题目：空间非合作目标的智能感知与控制

报告摘要：空间运动目标捕获在包括垃圾清理、空间站维护、非合作目标转移等空间在轨服务中发挥着重要的作用。捕获非合作空间运动目标的关键步骤包括运动目标感知、位姿估计和自主抓捕控制。运动目标感知利用视觉、激光雷达等多种传感器获取目标的基本信息并跟踪其运动,然后通过对运动目标的运动姿态和速度进行估计。然后通过有效的自主柔顺捕获控制以避免运动目标与航天器之间的刚性碰撞。近年来,哈尔滨工业大学自主智能无人系统工信部重点实验室对空间运动目标的感知、位姿估计与抓捕技术进行了深入的研究。本次报告将重点探讨空间目标高精度感知的快速角点提取策略、用于实时运动估计的点云配准数据融合新方法,以及用于任务空间跟踪和自主捕获的先进控制方案,并将讨论这些方法在实际应用中的关键问题和挑战。

个人简介：吴立刚, 1977 年生, 博士, 哈尔滨工业大学航天学院教授、长江学者特聘教授、国家杰出青年科学基金获得者、万人计划科技创新领军人才。目前担任哈尔滨工业大学航天学院院长、激光空间信息全国重点实验室主任、自主智能无人系统工信部重点实验室主任。主要从事非连续动态系统、自主智能无人系统、智能机器人技术、机器视觉与智能检测等领域研究。目前出版英文专著 9 部, 发表国际期刊论文 260 余篇, SCI 引用 20000 余次; 研究成果获黑龙江省自然科学一等奖 2 项 (第一完成人)、国家自然科学基金二等奖 1 项 (第二完成人)。牵头获国家重点研发计划、国家自然科学基金重点项目等 20 余项基金和项目资助; 获中国青年五四奖章、中国侨界贡献奖一等奖等荣誉。先后担任《IEEE 自动控制汇刊》、《IEEE 工业电子汇刊》、《自动化学报》等十余个期刊编委职务。2015 年以来连续当选全球高被引学者, 2020 年当选 IEEE Fellow, 2022 年当选中国自动化学会 (CAA) 会士。



特邀讲者 2：高连如 中国科学院研究员

报告题目：高光谱遥感图像目标探测技术

报告摘要：目标探测是高光谱遥感图像处理与信息提取中重要的研究和应用方向之一，为我国的民生工程及国防安全建设提供了重要技术支撑。本报告针对高光谱图像目标探测领域的相关研究进行综述，结合已有的成熟理论体系、各类经典的模型机理以及相关实验经验，从信号处理、机器学习和深度学习等三个方面进行梳理归纳。在此基础上，从高性能处理器件、算法实时化和在轨实时处理等方面，介绍面向遥感卫星平台的在轨实时高光谱图像目标探测技术。最后，以多学科交叉、软硬件结合以及理论与应用相互促进为切入点展望未来。

个人简介：高连如，中国科学院空天信息创新研究院研究员、博导，研究方向为高光谱遥感图像处理与信息提取，是 IET Fellow、APSIPA Distinguished Lecturer。主持国家自然科学基金委的杰青、优青项目以及国家重点研发计划课题、国家高分专项项目等，已发表学术论文 260 余篇，其中 SCI 收录 160 余篇，ESI 高被引论文 25 篇；授权国家发明专利 30 项；出版《高光谱图像信息提取》等学术著作 3 部。获得国家科技进步二等奖、中国科学院杰出科技成就奖、军队科技进步一等奖各 1 项；获得国际高光谱遥感顶级会议 WHISPERS 的杰出论文奖 1 项。现担任 IEEE TGRS、IET Image Processing 期刊的副主编以及 Chinese Geographical Science、《遥感学报》的编委，担任 IGARSS 国际会议评奖委员会委员和 WHISPERS 国际会议技术委员会委员。



特邀讲者 3: 李盛阳 中国科学院教授

报告题目: 空间对地观测视频智能处理-挑战、数据集与方法

报告摘要:空间对地观测视频卫星技术与星座构建迅猛发展，视频智能处理方法的研究与应用也越来越迫切。报告首先介绍了空间对地观测视频智能处理存在困难与挑战，重点报告了已制作和发布的空间对地观测视频多任务数据集包括场景分类、目标检测、目标跟踪、目标分割及超分辨率等任务，侧重阐述了多标签场景分类、有限监督条件下的目标检测、单目标和多目标跟踪等多任务方面取得的科研进展与成果，最后对该方向的研究进行了总结和展望。

个人简介:李盛阳，研究员、博导、中国科学院大学岗位教授，兼任中国图象图形学会遥感图象专业委员会委员、中国人工智能学会模式识别专业委员会委员、中国光学工程

学会空间信息技术专业委员会委员、中国载人航天工程软件专家组成员等。获省/部级科技奖 3 项，被国家表彰为“中国载人航天工程突出贡献者”，获国家发明专利 18 项，负责完成国家级标准 8 项，发表高水平学术论文一百余篇，指导研究生多次获得国内外有重要影响力人工智能方面竞赛冠军。



特邀讲者 4: 杨文 武汉大学电子信息学院教授

报告题目: 高分光学遥感图像小目标检测

报告摘要: 高分遥感影像小目标检测在国防安全、城市管理、环境保护和农业应用等多个领域具有重要的应用价值。然而, 精准检测高分辨率光学遥感图像中的小目标是一项极具挑战的任务。一方面, 现有遥感光学小目标检测数据集匮乏, 限制了相关算法研究及其在下游场景的应用; 另一方面, 小目标包含的像素数少、外观信息弱, 使得最先进的通用目标检测器无法在遥感小目标检测上获得令人满意的结果。本报告重点聚焦遥感小目标检测标注数据集匮乏、水平框遥感小目标检测训练中缺少正样本监督以及旋转框遥感小目标检测训练架构中的样本不平衡和特征不匹配等问题, 提出了相应的数据集、基准和算法, 并对下一步工作给出了相关建议。

个人简介: 杨文, 武汉大学电子信息学院教授, 博士生导师, 长期致力于高分遥感影像理解及其应用研究。主持和参与国家重点研发计划、国家高技术研究发展计划项目、国家自然科学基金等项目二十余项, 以第一作者或通讯作者发表 SCI 检索论文 80 余篇, 获批国家发明专利 20 余项。研究成果曾获湖北省科学技术奖励自然科学奖一等奖(2024)、二等奖(2018) 和三等奖(2012)、国际摄影测量与遥感杂志(ISPRS J P&RS)最佳论文奖(2021)、CVPR 最佳论文候选(2021)等。目前主要研究方向为高分影像目标检测与定位。

专题论坛：大模型赋能智慧医疗

（论坛主席：陈浩，香港科技大学计算机科学与工程系和化学与生物工程系助理教授，

庄吓海，复旦大学大数据学院教授，

徐军，南京信息工程大学人工智能学院副院长）

（10月20日 17:30-19:35，地点：报告厅）

时间	报告嘉宾姓名、单位、及 联系信息	报告题目
10月20日		
17:30-17:55	王满宁 复旦大学 mnwang@fudan.edu.cn	基于大模型的病理图像小样本学习
17:55-18:20	王连生 厦门大学 lswang@xmu.edu.cn	病理数据的多模态分析
18:20-18:45	张永兵 哈尔滨工业大学（深圳） ybzhang08@hit.edu.cn	全切片病理扫描成像与计算
18:45-19:10	余晋刚 华南理工大学 jgang.yu@gmail.com	迈向临床实用级组织病理学 AI 辅助诊断
19:10-19:35	雷柏英 深圳大学 leiby@szu.edu.cn	大模型赋能阿尔茨海默病智能诊断

论坛介绍

近年来，人工智能技术的飞速发展推动了医疗行业的智能化转型。其中，大模型凭借其强大的语义理解和生成能力，正在成为智慧医疗领域的关键技术之一。大模型是指通过对广泛数据进行预训练而能够适应多种任务的模型，能够帮助医疗机构实现医疗影像分析、辅助诊断、疾病发展趋势预测、个性化治疗方案制定等智能化应用，为临床决策提供重要支持，有助于打破有限的人工智能模型难以满足多样化医疗实践需求的矛盾，使更广泛的医疗场景受益于医疗大模型的发展，从而提升智慧医疗服务的水平。然而，尽管大模型在智慧医疗领域展现出巨大的潜力，其实际应用过程中仍面临许多挑战，例如模型的可解释性和透明度问题、患者隐私保护以及伦理合规等问题。为此，本次研讨会将邀请来自计算机、医疗、生物信息等多个交叉领域的行业专家，促进跨学科合作，分享基于大模型的智慧医疗研究成果和实践经验，共同探讨大模型在智慧医疗中的最新进展和未来趋势。

论坛主席



个人简介: 陈浩(jhc@ust.hk), 香港科技大学计算机科学与工程系和化学与生物工程系助理教授, 研究兴趣包括可信人工智能, 医学图像分析, 深度学习等。他领导的人工智能医疗实验室 (Smart Lab), 专注于可信人工智能技术在医疗领域的前沿研究与转化应用。陈教授于2017年获得香港中文大学博士学位。在 MICCAI、IEEE-TMI、MIA、CVPR、ICCV、AAAI、IJCAI、Radiology、Nature Communications、Lancet Digital Health、Nature Machine Intelligence 等顶级期刊和会议发表论文 200 余篇 (谷歌学术引用次数 26900 余次, h-index 65), 入选 2022 年和 2023 年斯坦福大学全球排名前 2% 科学家名单。此外, 陈教授还具有丰富的工业研究和产业转化经验, 拥有二十余项人工智能和图像分析方面专利。曾获得 2023 年亚洲青年科学家、国家教育部高等学校科学研究优秀成果二等奖、北京市科技进步一等奖、2019 年人工智能医学影像顶级会议 MICCAI 青年科学家影响力奖、Elsevier-MICCAI 最佳论文奖、医学影像与增强现实会议最佳论文奖、福布斯中国 30 岁以下 30 位精英等奖项, 担任包括 IEEE TMI、TNNLS、J-BHI、CMIG 和 Medical Physics 等期刊编委, 担任 2025 ICLR、2024 CVPR、2024 ACM MM、2021-2023 MICCAI、2022-2024 MIDL 等多个人工智能与医学影像分析国际会议的领域主席和程序委员, 曾带领团队获得 15 余项国际医学图像分析的挑战赛冠军。



个人简介: 庄吓海(zxh@fudan.edu.cn), 复旦大学大数据学院教授、博导。主要研究方向是心脏医学影像分析, 可解释人工智能方法等。在本领域主流期刊发表文章百余篇; 多项工作入选顶刊 Med Imag Ana 高引论文, 工作被 Nat Rev Card 2021 作为人工智能在心血管成像中的成功应用案例引用, 以及 IEEE T PAMI、Circulation 和多个 Nature 子刊等引用。2022 年入选斯坦福大学发布全球顶尖科学家“终身”和“年度”科学影响力榜单; 曾提名 MICCAI 协会青年科学家奖。担任国际学术组织 MICCAI 协会的理事, 担任包括三个中科院一区期刊等多个国际期刊的编委。



个人简介: 徐军(xujunus@gmail.com), 南京信息工程大学人工智能学院副院长, 二级教授, 智慧医疗研究院执行院长。2007 年于浙江大学控制科学与工程系获博士学位, 先后在美国 Rutgers 大学、凯斯西储大学生物医学工程系任博士后研究员和访问教授。兼任江苏省人工智能学会医学图像分析专委会副主任委员; 中华医学会病理学分会数字病理与人工智能委员会委员; 第四届医学图像计算青年研讨会轮值主席等。主要研究成果发表在 Nature Communications, Radiology, IEEE Trans. on Medical Imaging, Medical Image Analysis 等期刊。入选斯坦福大学全球前 2% 顶尖科学家榜单。先后主持国家自然科学基金联合重点项目、面上项目、科技部重点研发计划重点专项课题、省部级项目等多项。参与国家自然科学基金重大研究计划“肿瘤演进与诊疗的分子功能可视化研究”重点项目和重大集成项目多项。主要研究方向包括医学图像计算、计算病理、基于影像和常规病理切片定量分析的疾病辅助预防、诊疗和预后。致力疾病的 1) 自动识别与诊断及亚型的分类, 2) 影像和病理图像亚视觉特征的量化和表征, 3) 转移风险、恶性程度、治疗响应的预测、预后, 期望这些研究成果能够辅助医生进一步改善疾病的预防、诊疗和预后, 让患者获益。

特邀讲者简介



特邀讲者 1：王满宁 复旦大学教授

报告题目：“基于大模型的病理图像小样本学习”

报告摘要：病理诊断是众多疾病临床诊断的金标准，数字病理图像的智能诊断不但可以帮助医生提高诊断的效率和准确性，而且可以识别人类肉眼无法识别的图像特征，完成一些病理学家无法完成的诊断任务，因此具有重要的临床价值。近年来出现了多种数字病理大模型，并在多个下游任务上表现出良好的性能。但是，由于病理诊断的应用场景多种多样，大模型难以在更多任务上实现零样本学习，因此利用大模型开展病理智能诊断的小样本学习研究具有重要的临床价值。本报告将介绍课题组近年来在数字病理图像小样本学习方面的研究工作，主要介绍如何利用图像-文本大模型实现全切片级别的病理智能诊断。

个人简介：王满宁(mnwang@fudan.edu.cn)，复旦大学数字医学研究中心教授、博士生导师。长期从事医学图像处理与手术导航领域的研究工作，主要研究方向包括：三维点云空间配准、数字病理智能诊断、多模态医学影像智能诊断、主动学习与域适应等。先后在 IEEE TPAMI、IEEE TIP、IEEE TMI、NeurIPS、CVPR、ICCV、ECCV 等杂志和会议上发表研究论文 100 余篇；获得国家技术发明二等奖（第 2 完成人）、上海青年科技英才、上海市优秀技术带头人等奖励。担任中国计算机学会数字医学分会常委、中国自动化学会模式识别与机器智能专委会委员、上海市医学会数字医学专科分会副主任委员、上海市生物医学工程学会医学人工智能专委会委员、医学图像计算青年研讨会（MICS）委员。



特邀讲者 2: 王连生 厦门大学教授

报告题目: “病理数据的多模态分析”

报告摘要: 应用人工智能技术对病理数据进行分析, 已经取得了显著进展。以病理数据为基础, 整合临床信息、基因组学数据、影像学资料及生物标志物等多维度临床医疗信息, 探索多模态数据在提升疾病诊断准确性、个性化治疗方案设计及病情进展评估等方面的潜力与挑战, 将为精准医疗发展提供新视角。

个人简介: 王连生(lswang@xmu.edu.cn), 博士毕业于香港中文大学, 现为厦门大学信息学院教授, 医学院消化病学系双聘教授, 博士生导师, 数字福建健康医疗大数据研究所副所长, 福建省医学会放射学分会 AI 学组副组长, 厦门大学医学人工智能研究院(筹)负责人, MICS 副主席。长期从事医学影像处理研究, 主持和参与多项科研项目, 包括国家自然科学基金仪器专项、科技部科技创新 2030、国家重点研发项目、国家自然科学基金面上和青年项目等, 发表包括 Nature Machine Intelligence、Nature Communications、Cell Patterns、TMI、MedIA、人工智能顶会 CVPR/AAAI 等相关研究论文 100 余篇, 获得腾讯犀牛鸟科研奖、CSPE Young Investigator、福建省科技进步二等奖、“CCF-腾讯犀牛鸟拔尖人才培养卓越导师”奖、2023 年厦门大学田昭武交叉学科奖一等奖, 带领团队先后 11 次在国际医学影像比赛中获得冠军。



特邀讲者 3：张永兵 哈尔滨工业大学（深圳）教授

报告题目：“全切片病理扫描成像与计算”

报告摘要：基于病理成像与计算的组织学分析是临床上进行癌症诊断的黄金标准。然而，在成像方面，现有方法存在成像速度慢、精度低等局限；在病理诊断方面现有方法存在人工阅片诊断工作量大、受主观因素限制等缺点，容易造成误诊漏诊等缺陷。如何从理论上突破多方面局限，精准高效地快速获取高分辨病理切片图像，进而智能实时地辅助病理医生做出准确的疾病诊断和预后，成为病理组织成像和计算中亟待解决的问题。本次报告首先介绍了计算病理的发展历程；接着分别介绍了课题组在快速全切片病理扫描成像与人工智能病理计算方面的研究工作，最后对病理成像与计算的发展给出了展望。

个人简介：张永兵(ybzhang08@hit.edu.cn)，哈尔滨工业大学（深圳）计算机学院教授、博士生导师。主要研究领域为计算机视觉、生物医学图像处理、计算成像等。为中国计算机学会、中国人工智能学会、IEEE、SPIE、OSA 等多个国内外知名学会会员。2010年在哈尔滨工业大学取得博士学位。2010-2012年，在清华大学深圳国际研究生院进行博士后研究，2012年留校任教，并于2014年晋升为副研究员，期间曾作为访问学者赴美国加州大学伯克利分校从事访问研究。自2021年起，加入哈尔滨工业大学（深圳）计算机学院，担任教授，继续从事教学和科研工作。研究方向为计算机视觉、生物医学图像处理、计算成像等，在相关领域发表论文100余篇，多篇文章发表在人工智能和计算机视觉顶会NeurIPS、AAAI、CVPR、ICCV、TIP、MICCAI等，获国际会议论文奖2项；授权发明专利50余项，美国发明专利2项。目前主要研究工作是探索人工智能与计算机视觉在生命医学和医疗健康领域的应用。



特邀讲者 4：余晋刚 华南理工大学副教授

报告题目：“迈向临床实用级组织病理学 AI 辅助诊断”

报告摘要:过去一年，全国多地陆续出台新政加快推进数智化病理建设，病理大模型不断涌现，病理 AI 正迎来新一轮发展机遇。然而，受病理图像数据的长尾分布、异质性、标注困难、标准化缺失等特性的影响，现有病理 AI 算法的性能通常无法达到临床实用级水平，从而难以真正部署应用于临床。本报告结合我们组织病理学多病种 AI 辅助诊断、免疫组织化学 AI 判读这两个典型应用场景相关的基础算法研究与转化应用实践，分析探讨构建一个临床实用级病理 AI 算法模型所面临的关键技术挑战及可行解决方案。

个人简介: 余晋刚(jgang.yu@gmail.com)，华南理工大学自动化学院副教授、博士生导师，广东省“珠江人才计划”青年拔尖人才入选者。先后在西安交通大学、华中科技大学、美国内布拉斯加林肯大学完成学业及博士后训练。长期从事医学图像分析领域的研究工作，近年来主要聚焦于病理人工智能相关的基础算法与转化应用研究。以第一或通讯作者在 TMI、MedIA、MICCAI、TIP、CVPR 等领域高水平期刊和会议发表论文 30 余篇；主持国家自然科学基金项目（2 项）、国家重点研发计划子课题、省级人才计划项目和一般项目（4 项）、校企产学研合作项目等 10 余项；授权国家发明专利 8 件，已转化金额近 500 万元，主持研发的组织病理学 AI 辅助诊断产品已在多家医院落地应用。担任琶洲实验室双聘研究人员、南方医科大学珠江医院流动 PI、中国图象图形学会视觉认知与计算专委会副秘书长、医学图像计算青年研讨会（MICS）委员、CVPR/ICCV/ECCV/IJCAI/AAAI 会议程序委员等学术兼职。



特邀讲者 5：雷柏英 深圳大学教授

报告题目：“大模型赋能阿尔茨海默病智能诊断”

报告摘要：人工智能技术在人体健康诊断方面取得了显著的进展，尤其是基于深度学习的大模型（Large-scale Model），如 BERT、GPT 等，展现出了强大的数据驱动和自我学习能力。针对阿尔茨海默病（AD）智能诊断难以满足临床需求的困境，基于大模型的新型研究思路有望促进该研究的发展。大模型主要基于 Transformer 网络架构，此外涉及脑网络构建、早期生物标志物发掘、多模态多组学数据融合、多中心学习等多方面的研究基础。因此我们将对该研究思路的相关研究基础进行详细介绍，并对大模型在 AD 智能诊断中的应用进行展望。

个人简介：雷柏英(leiby@szu.edu.cn)，国家级青年人才，深圳大学特聘教授，西安电子科技大学客座教授，博士生导师，深圳市海外高层次人才（孔雀计划）、获新加坡南洋理工大学博士学位。主要研究方向为医学图像处理 and 人工智能。在 IEEE TPAMI、TMI 等以第一/通讯作者（含共同）发表 SCI 论文 100 余篇（6 篇 ESI 高被引，1 篇热点论文）。谷歌学术总引用超万次，H 指数 51。主持国家自然科学基金联合基金重点 1 项，面上 2 项等项目 20 余项（含国家级 7 项）。现任 IEEE TNLS、TCYB、TMI、JBHI、Medical Image Analysis 等 10 种 SCI 期刊编委。IEEE BISP、BIIP、BSP、BHI 等技术委员会委员，医学图像顶级学术会议 MICCAI 领域主席（2021-2023）。IEEE 广州分部 WIE 主席，人工智能 A 类会议 AAAI、IJCAI 程序委员会委员，入选美国斯坦福大学发布的“全球前 2% 顶尖科学家”（2020-2023），获“强国青年科学家”提名（2022，全国共 40 人），CSIG 石青云女科学家奖（2022）。

讲习班信息与讲者简介

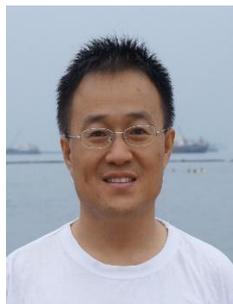
场景感知：从生物视觉机理到类脑视觉模型

（组织者/主持人：李永杰 电子科技大学）

（讲者：邢大军 北京师范大学；杨开富 电子科技大学）

（10月18日 10:00—11:45，地点：报告厅）

讲习班组织者/主持人简介



李永杰，电子科技大学生命科学与技术学院、神经信息教育部重点实验室教授、博士生导师、“视觉认知与类脑计算”团队负责人，教育部新世纪优秀人才（2007），兼任中国自动化学会理事及生物控制论与生物医学工程专委会主任委员、中国电子学会生物医学电子学分会副主任委员、中国人工智能学会模式识别专业委员会委员、中国计算机学会计算机视觉专业委员会委员等。2004年在电子科技大学取得生物医学工程专业博士学位。2009.9-2010.10在哥伦比亚大学神经科学系做访问学者。主要研究方向为大脑视觉感知计算模型及计算机视觉应用。先后在IEEE Trans on PAMI/IP/MI/CSVT、IJCV、PR、NeuroImage等期刊及ICCV/CVPR/ECCV/ACMMM等会议上发表论文百余篇，申请中国发明专利40余项（已授权30余项），曾获吴文俊人工智能自然科学奖三等奖（2020）。

（email: liyj@uestc.edu.cn）

讲习班讲者简介

邢大军 (报告主题: 生物视觉感知机理)



邢大军, 北京师范大学认知神经科学与学习国家重点实验室、IDG/McGovern Institute for Brain Research 教授、研究员, 国家中组部“青年千人计划”(2014), 教育部新世纪优秀人才(2014), 北京神经科学学会理事。2005 年获纽约大学神经生物学博士学位。2005-2013 年先后任纽约大学神经科学中心任研究学者, 纽约市立大学亨特学院心理学系任兼职助理教授以及美国纽约大学神经科学中心任助理研究教授。主要的研究方向为利用电生理、数学建模、信号处理、心理物理以及解剖学等手段, 系统地研究高等哺乳动物(猕猴和人)的大脑视觉信息处理过程中, 群体编码、信息加工以及脑部振荡信号的神经机理和数学模型。相关工作发表在 Nature Communications、PNAS、PLoS Biology、Cell Reports 和 Journal of Neuroscience 等期刊。

(email: dajun_xing@bnu.edu.cn)

杨开富 (报告主题: 视觉计算模型与应用)



杨开富, 电子科技大学生命科学与技术学院、神经信息教育部重点实验室副研究员, 硕士生导师, 中国人工智能学会模式识别专业委员会委员、中国图像图形学学会视觉认知与计算专业委员会委员。2016 年在电子科技大学取得生物医学工程专业博士学位并留校任副研究员。2019.08-2020.08 在苏黎世联邦理工学院(ETH Zurich)计算机视觉实验室做访问学者。主要研究方向为视觉感知与认知计算、类脑图像处理与计算机视觉应用等。已在 IEEE Trans. PAMI/IP/CSVT, IJCV 等期刊及 ICCV/CVPR/ECCV 等国际会议上发表学术论文 40 余篇, 申请中国发明专利 10 余项。曾获 CSIG 优秀博士论文提名奖(2018)、吴文俊人工智能自然科学奖三等奖(2020)、电子科技大学学术新人奖(2022)。

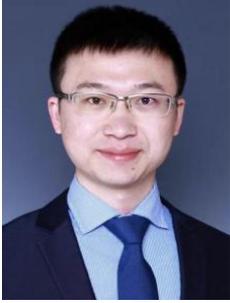
(email: yangkf@uestc.edu.cn)

多模态基础模型的结构设计与优化

(组织者/主持人: 黄高 清华大学; 刘静 中国科学院自动化研究所; 杨乐 西安交通大学)

(讲者: 周奕毅 厦门大学; 郭龙腾 中国科学院自动化研究所; 王文海 香港中文大学)
(10月18日 10:00—11:45, 地点: 塔里木厅)

讲习班组织者/主持人简介



黄高, 清华大学自动化系副教授, 博士生导师。博士毕业于清华大学, 博士后工作于美国康奈尔大学。主要研究领域为深度学习和计算机视觉, 提出了主流卷积网络模型 DenseNet。共计发表学术论文 100 余篇, 被引 6 万余次, 最高单篇引用超过 4 万次。获国家优青、CVPR 最佳论文奖、达摩院青橙奖、世界人工智能大会 SAIL 奖、教育部自然科学一等奖、AI 2000 人工智能最具影响力学者、《麻省理工科技评论》

“35 岁以下科技创新 35 人”等, 担任 IEEE-TPAMI、IEEE-TBD、Pattern Recognition 等国际期刊编委和 CVPR、ICCV、NIPS、ICML 等国际会议领域主席。



刘静, 中科院自动化所研究员, 博士生导师, 中国科学院大学岗位教授, 国家优青获得者。研究方向图像语义理解、视觉与语言、多模态预训练等。提出国际首个图文音三模态“紫东太初”大模型, 显著提升多模态理解与生成任务的准确性与泛化性。已发表高水平学术论文 150 余篇, 谷歌学术引用 7700+次, SCI 他引次数 2500+次, 其中有三篇被 ESI 列为 Top 1% 高被引论文。相关研究成果在华为、京东等产品中得到实际应用, 曾获图像图形学会科学技术二等奖。相关领域的

国际学术竞赛中荣获 10 项冠军。



杨乐, 西安交通大学电信学部, 信息与通信工程学院助理教授, 博士毕业于清华大学自动化系, 获 2021 年“博新计划”, 合作导师为管晓宏院士。主要研究方向为高效神经网络结构设计与计算机视觉。在 CVPR、IEEE TNNLS、IEEE TCYB 等高水平期刊/会议发表论文 30 余篇, 谷歌学术引用 1500 余次。主持国家自然科学基金青年基金, 新一代人工智能国家科技重大专项子课题, CCF-百度松果基金等项目。担任 IEEE TPAMI、IJCV、CVPR、ICML、ICLR 等期刊/会议审

稿人及 IJCAI 会议程序委员会成员。

讲习班讲者简介

周奕毅（报告主题：机遇与挑战—多模态语言大模型研究进展介绍）



周奕毅，厦门大学信息学院副教授，博士生导师、厦门市“双百计划”领军人才、厦门大学南强青年拔尖人才。2019年博士毕业于厦门大学，博士生导师为纪荣嵘教授。研究方向为视觉-语言多模态学习，在IEEE TPAMI、IJCV、IEEE TIP、CVPR、ICCV、NeurIPS等CCF推荐期刊及会议发表论文40余篇，多篇论文被评选为CVPR, ACM Multimedia和ECCV会议口头报告文章（Oral Paper）。主持某科委1XX基金、国家自然科学基金青年基金、博士后特别资助、厦门大学校长基金揭榜挂帅及华为校企合作等项目，并作为核心成员和执行负责人参加科技创新2030、某科技委基础加强计划和GF创新特区等多个人工智能重大项目。担任CCF YOSCEF厦门分论坛2024-2025主席，ACM Multimedia Meta Reviewer、AAAI Senior PC、PRCV 2023赞助主席，TIP、TMM、TNLS、PR等期刊审稿人及CVPR、ICCV、ECCV、NeurIPS等国际重要学术会议程序委员会委员。

郭龙腾（报告主题：多模态预训练模型的构建与应用）



郭龙腾，中国科学院自动化研究所副研究员，紫东太初大模型研究中心骨干成员。研究方向包括多模态预训练模型、多模态理解与生成等。曾获中科院院长奖、2022年世界人工智能大会“卓越人工智能引领者奖SAIL”。承担或参与多项国家自然科学基金项目、2030新一代人工智能重大项目等。在相关领域发表高水平学术论文二十余篇，在图像、视频理解相关的国际学术竞赛中获得冠军4+项。

王文海（报告主题：视觉与语言的交汇—图文多模态大模型的研究与实践）



王文海，南京大学博士，香港中文大学博士后。研究方向为视觉基础模型研究，上海人工智能实验室“书生”系列视觉基础模型核心开发者。主要成果发表在顶级期刊和会议TPAMI、CVPR、ICCV、ECCV、ICLR、NeurIPS等共43篇论文，其中19篇为一作/共一/通信。研究成果获得了总共超1.5万次引用，单篇最高引用超3000次。研究成果分别入选CVPR 2023最佳论文，世界人工智能大会青年优秀论文奖，CVMJ 2022最佳论文提名奖，两次入选ESI高被引论文（前1%）和热点论文（前0.1%），6次入选Paper Digest CVPR、ICCV、NeurIPS、ECCV年度十大最具影响力论文，入选2023年CSIG优博提名，斯坦福大学2023年度全球前2%顶尖科学家。

医疗大模型

(组织者/主持人: 雷柏英 深圳大学; 胡众义 温州大学)

(讲者: 何晖光 中国科学院自动化研究所; 陈浩 香港科技大学;
陈俊颖 华南理工大学; 杨二昆 西安电子科技大学 房钰棋 南京大学)

(10月18日 12:00—14:00, 地点: 学术厅)

讲习班组织者/主持人简介



雷柏英, 国家级青年人才, 深圳大学特聘教授, 西安电子科技大学客座教授, 博士生导师, 深圳市海外高层次人才(孔雀计划)、获新加坡南洋理工大学博士学位。主要研究方向为医学图像处理 and 人工智能。在 IEEE TPAMI、TMI 等以第一/通讯作者(含共同)发表 SCI 论文 100 余篇(6 篇 ESI 高被引, 1 篇热点论文)。谷歌学术总引用超万次, H 指数 51。主持国家自然科学基金联合基金重点 1 项, 面上 2 项等项目 20 余项(含国家级 7 项)。现任 IEEE TNNLS、TCYB、TMI、JBHI、Medical Image Analysis 等 10 种 SCI 期刊编委。IEEE BISP、BIIP、BSP、BHI 等技术委员会委员, 医学图像顶级学术会议 MICCAI 领域主席(2021-2023)。IEEE 广州分部 WIE 主席, 人工智能 A 类会议 AAAI、IJCAI 程序委员会委员, 入选美国斯坦福大学发布的“全球前 2% 顶尖科学家”(2020-2023), 获“强国青年科学家”提名(2022, 全国共 40 人), CSIG 石青云女科学家奖(2022)。(email: leiby@szu.edu.cn)



胡众义, 博士, 教授。温州大学智能信息系统研究所负责人, 温州市智能影像处理与分析重点实验室主任。中国人工智能学会智能空天专委会常务委员, IEEE SENIOR, 中国计算机学会高级会员, 中国人工智能学会终身会员, 中国自动化学会终身会员, 温州知识产权法庭兼职技术调查官等。长期从事计算机与人工智能应用教学及研究, 感兴趣的方向包括医学图像计算、机器视觉、机器学习、智能信息处理等。迄今为止已在国内外学术期刊和会议上发表论文 50 余篇, 申请专利 20 余项, 登记计算机软件著作权 10 余项, 出版图像处理著作 1 部。主持国家自然科学基金联合重点支持项目课题(合作单位负责人)、浙江省自然科学基金重大项目、重点项目等 20 余项。(email: huzhongyi@wzu.edu.cn)

讲习班讲者简介

何晖光（报告主题：多模态大模型在脑信息编解码中的应用）



何晖光，中国科学院自动化研究所研究员，博士生导师，中国科学院大学岗位教授，国家高层次领军人才。先后主持包括多项国家自然科学基金、863 项目、国家重点研究计划课题等多个重要项目。获得国家科技进步二等奖两项（排二、排三），北京市科技进步奖两项，教育部科技进步一等奖，中科院首届优秀博士论文奖等奖项。入选北京市科技新星，中科院“卢嘉锡青年人才奖”，中科院青年创新促进会优秀会员等。其研究领域为人工智能，脑-机接口、医学影像分析等，在 IEEE TPAMI, ICML 等发表文章 200 余篇。CCF/CSIG 杰出会员。建国七十周年纪念章获得者。

陈浩（报告主题：大模型赋能计算病理）



陈浩，香港科技大学计算机科学与工程系和化学与生物工程系助理教授，研究兴趣包括人工智能，医疗图像分析，深度学习等。他领导的人工智能医疗实验室（Smart Lab），专注于可信赖人工智能技术在医疗领域的前沿研究与转化应用。陈博士于 2017 年获得香港中文大学博士学位。在 MICCAI、IEEE-TMI、MIA、CVPR、ICCV、AAAI、IJCAI、Radiology、Lancet Digital Health、Nature Machine Intelligence、JAMA 等顶级期刊和会议发表论文 100 余篇（谷歌学术引用次数达到 27100 余次，h-index 65），入选 2022 年和 2023 年斯坦福大学全球排名前 2% 科学家名单。此外，陈博士还具有丰富的工业研究和产业转化经验，拥有二十余项人工智能和图像分析方面专利。曾获得 2023 年亚洲青年科学家、国家教育部高等学校科学研究优秀成果二等奖、北京市科技进步一等奖、世界人工智能大会最高奖项卓越人工智能引领者(SAIL)奖、2019 年人工智能医学影像顶级会议 MICCAI 青年科学家影响力奖、Elsevier-MICCAI 最佳论文奖、医学影像与增强现实会议最佳论文奖、福布斯中国 30 岁以下 30 位精英、香港资讯及通讯科技银奖等奖项，担任包括 IEEE TMI、TNNLS、J-BHI、CMIG 和 Medical Physics 等期刊编委，担任 ICLR 2025、CVPR 2024-2025 ACM Multimedia 2024、MICCAI 2021-2023 等多个人工智能与医学影像分析国际会议的领导主席和程序委员，曾带领团队获得 15 余项国际医学图像分析的挑战赛冠军。（email: jhc@cse.ust.hk）

陈俊颖（报告主题：大模型赋能临床疾病诊断）



陈俊颖，华南理工大学教授、博士生导师，国家卓越工程师学院导师，大数据与智能机器人教育部重点实验室主任助理、智能医学图像处理与医疗机器人研究室负责人，香港大学博士，浙江大学学士（优秀毕业生）。主要研究多源多模态特征融合、医学图像处理与成像、智能机器人等，成果发表于 TPAMI、TNNLS、TMI、CVPR、SMC、IROS 等期刊会议，获得 2023 年广东省计算机学会青年科技奖、2023 年广东省人工智能产业协会青年科技创新奖、2023 年广东省计算机学会科学技术奖科技进步一等奖（第一完成人）、2023 年广东省精准医学科学技术奖科技创新一等奖（第一完成人）、2018 年中国计算机学会科学技术奖技术发明二等奖等奖励，获评国际计算机学会（ACM）特色杰出演讲者、中国计算机学会杰出演讲者、中国计算机学会杰出会员等荣誉称号。（email: jychense@scut.edu.cn）

杨二昆（报告主题：噪声标签学习及其在医学和大模型中的应用）



杨二昆，西安电子科技大学副教授，省级青年人才。分别于 2013 年、2018 年从西安电子科技大学获得信息对抗技术专业学士学位和电路与系统专业博士学位，2017-2019 年在澳大利亚悉尼大学进行博士生联合培养，2019-2021 年在美国北卡罗莱纳大学教堂山分校做博士后研究。2021 年回校工作，隶属于邓成教授领导的 OPTIC 实验室。主要研究方向为机器学习，计算机视觉，多媒体信息检索，噪声标签学习等，已在国际顶级期刊或会议上发表论文四十余篇，担任 IEEE T-PAMI, IEEE T-NNLS, IEEE T-IP 等十余种国际顶级期刊审稿人，及 CVPR, NeurIPS, ICLR, ICCV, KDD, AAAI, IJCAI, MICCAI 等十余种国际顶级会议的（高级）程序委员会委员，获陕西省优秀博士学位论文，入选中国科协青年人才托举，获陕西省自然科学一等奖（2/5）。

房钰棋（报告主题：大模型赋能影像智能诊疗技术）



房钰棋，南京大学智能科学与技术学院助理教授，博导。2020 年于香港中文大学获博士学位，2021 年 9 月至 2024 年 4 月于北卡罗莱纳大学教堂山分校从事博士后研究。主要研究方向为基于人工智能和医学数据分析的跨学科领域，旨在创建智能医疗系统推进临床决策及医疗保健，包括疾病诊断和病灶检测等。在 Medical Image Analysis、MICCAI 等领域一流国际期刊和会议上发表高水平论文十余篇，获批发明专利两项。长期担任 TPAMI、MedIA、TMI、PR、MICCAI 等国际权威期刊及会议审稿人。（email: yqfang@nju.edu.cn）

基于深度学习的智能疾病诊断与个性化手术规划

(组织者/主持人: 陈梁骏 西安交通大学)

(讲者: 吴焯 南京理工大学; 杨鹏 深圳大学; 武佳懿 西安交通大学;
马磊 同济大学; 张栋 西安交通大学)

(10月18日 12:00—14:00, 地点: 塔里木厅)

讲习班组织者/主持人简介



陈梁骏, 西安交通大学人工智能学院特聘研究员, 获陕西省“三秦英才引进计划”和西安交通大学“青年拔尖人才计划”等项目支持。分别于2012年6月和2019年7月获西安交通大学本科和博士学位, 2019年7月至2023年3月于北卡罗莱纳大学教堂山分校师从沈定刚教授做博士后研究。2023年5月至今入职以郑南宁院士为学术带头人的西安交通大学人工智能与机器人研究所。主要研究方向为医学图像处理、人工智能和脑机接口, 发表领域内高水平论文20余篇,

包括 Nature Communications、Cell Report、Medical Image Analysis 为代表的一作 Top 期刊论文 7 篇, 领域顶会 MICCAI 会议论文 2 篇, 多项成果已公开并得到广泛下载和应用。主持国科金面上项目 1 项, 参与科技部重点研发项目 1 项、国科金天重点项目 1 项、国科金天元交叉重点项目 1 项, 学术任职包括中国自动化学会智能健康与生物信息专委会委员、中国自动化学会机器人专委会委员。(email: liangjunchen@xjtu.edu.cn)

讲习班讲者简介

吴焯 (报告主题: 轻量化弥散磁共振成像技术)



吴焯, 南京理工大学/计算机科学与工程学院/高维信息智能感知与系统教育部重点实验室, 教授、博士生导师, 入选国家海外高层次青年人才计划、江苏特聘教授、南京市高层次人才计划、紫金青年学者。从事脑科学与人工智能交叉研究。曾先后在哈佛大学联合培养、在北卡罗莱纳州大学完成博士后研究。近五年在 Nature Methods, Pattern, NeuroImage, Medical Image Analysis, IEEE TMI, MICCAI, IPMI 等发表学术论文 50 余篇, 获 Medical Image Analysis 最佳论文奖、ISMRM 卓越论文奖、OHBM

杰出论文奖、MICCAI-NIH 奖、国际“互联网+”全国金奖、中国发明协会创新创业一等奖等。目前主持国家自然科学基金、国家高层次人才项目、国家重点研发计划子任务、江苏省重大人才工程项目、中央高校科研基金、科技部政府间重点专项子任务、南京市留学人才科技创新基金等。

杨鹏（报告主题：基于磁共振影像的阿尔茨海默病和强迫症智能诊断）



杨鹏，博士，深圳大学医学部生物医学工程学院助理教授，特聘副研究员，硕士生导师。美国宾夕法尼亚大学访问学者、深圳市优秀博士后，医学图像计算青年研讨会（MICS）委员会委员。主要研究方向：医学图像计算-脑疾病智能诊断，如阿尔茨海默病、强迫症等。研究成果主要发表在

Medical Image Analysis, IEEE Transactions on Medical Imaging, IEEE Transactions on Cybernetics, International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI) 等国际顶级 SCI 期刊和会议。主持国家自然科学基金青年项目，中国博士后基金面上项目，广东省基础与应用基础研究基金区域联合基金等，现为 IEEE TNNLS, MICCAI, ISBI 等国际期刊和会议的审稿人。

武佳懿（报告主题：临床需求驱动的医学影像人工智能）



武佳懿，西安交通大学人工智能学院助理教授，分别于 2014 年 9 月至 2015 年 3 月、2017 年 1 月至 2018 年 2 月公派赴香港理工大学和美国华盛顿大学西雅图分校联合培养学习，主要从事糖尿病性视网膜疾病的智能诊断和颈动脉粥样硬化的智能诊断等科研工作。主持国家重点研发项目子课题、陕西省重点研发子课题和西安交通大学学科交叉与合作等项目/课题，参与国家自然科学基金面上项目、陕西省青年创新团队项目等科研工作。先后发表学术论文 25 篇，其中国际权威期刊 Medical Physics 等 SCI 期刊论文 7 篇、中文期刊 1 篇，会议论文 17 篇（EI 收录 15 篇），获国家发明专利授权 7 项，并申请国家专利 4 项。长期受邀担任 JBHI 国际期刊和 ISBI 国际会议审稿人。

马磊（报告主题：深度学习驱动的个性化正颌手术规划）



杨马磊 (Lei Ma)，控制系助理教授，硕士生导师，上海市海外高层次人才。2018 年博士毕业于日本东京大学工学院。随后赴美国 University of North Carolina at Chapel Hill 进行博士后研究。2023 年至今任同济大学电子与信息学院控制科学与工程系助理教授。研究方向包括：口腔智能化诊断与治疗、智能医学图像处理、软硬组织建模、医疗机器人以及三维智能感知等。目前主持国家自然科学基金青年项目 1 项，参与国家自然科学基金面上、国家科技支撑技术项目、美国 NIH01 项目等多项国内国外项目。目前已在本领域已发表学术论文 40 余篇，包括 IEEE Transactions on Medical Imaging、Medical Image Analysis、IEEE Journal of Biomedical and Health Informatics、MICCAI 等本领域顶刊顶会。长期担任 IEEE Transactions on Medical Imaging(TMI), IEEE Transactions on Bioengineering (TBME), Medical Image Analysis(Media), IEEE Journal of Biomedical and Health Informatics (JBHI), Pattern Recognition(PR), MICCAI, IPCAI 等顶级期刊与会议审稿人。

张栋（报告主题：超声影像的可解释性辅助诊断方法）



张栋，西安交通大学助理教授，医工交叉博士后 A 类。于 2024 年从西安交通大学获得控制科学与工程博士学位。2024 年进入西安交通大学人工智能学院与西北医院联合医工交叉博士后流动站工作，主要研究方向为计算机视觉与医学图像处理，研究内容聚焦于基于人机混合增强智能的医学影像可解释性诊断方法等，目前已发表学术论文 30 余篇，以第一作者或学生一作发表论文 11 篇，包括 PR 2023、IEEE JBHI 2023、EUP 2022、FIM 2022、MP 2022 & 2020 等；申请国家发明专利 11 项，已授权 5 项；获批软件著作权 3 项；曾以第一完成人获国家级竞赛奖 9 项。主持中央高校基本科研业务费项目 1 项，作为核心骨干成员参与科技创新 2030—“新一代人工智能”重大项目课题、国家重点研发计划“变革性技术关键科学问题”重点专项子课题、省重点研发计划高校联合项目等人工智能与医学领域的项目 9 项。

AIGC 在生物特征识别与安全中的应用

(组织者/主持人: 孙哲南 中国科学院自动化研究所)

(讲者: 朱翔昱 中国科学院自动化研究所; 彭勃 中国科学院自动化研究所; 李琦 中国科学院自动化研究所; 赵健 西北工业大学)

(10月18日 15:30—17:15, 地点: 多功能厅)

讲习班组织者/主持人简介



孙哲南 (znsun@nlpr.ia.ac.cn), 中国科学院自动化研究所、多模态人工智能全国重点实验室研究员、博导, 中国科学院大学人工智能学院教授。长期聚焦在虹膜识别、人脸识别等生物特征识别前沿科技研究, 研究成果获得国家技术发明二等奖、中国图象图形学学会自然科学一等奖、北京市科技进步二等奖、中国专利优秀奖, 入选国家万人计划科技创新领军人才和科技部中青年科技创新领军人才, 在生物特征识别和模式识别、计算机视觉领域共发表 318 篇(部)论著, 其中国际期刊论文 95 篇, IEEE Transactions 论文 55 篇(包括 12 篇 IEEE TPAMI), 近 5 年 SCI 他引 1,209 次, 其中高被引论文 5 篇, Google Scholar 引用 12,445 次, H-index 指数 62; 获得国家发明专利授权 55 项。曾担任国际模式识别学会生物特征识别技术委员会主席 (Chair of IAPR Technical Committee on Biometrics), 当选 IAPR Fellow, 国际生物识别权威期刊 IEEE Transactions on Biometrics, Behavior, and Identity Science 编委。发表于《中国图象图形学报》的《生物特征识别学科发展报告》成为 2021 年度图象图形类引用最高的综述论文。

讲习班讲者简介

朱翔昱 (xiangyu.zhu@nlpr.ia.ac.cn)



朱翔昱, 博士, 副研究员, 2017 年至今在中国科学院自动化研究所工作, 从事生物特征识别和人工智能基础理论研究与应用。国际模式识别协会 (IAPR) 生物特征识别青年学者奖 (YBIA) 获得者 (每两年从全球范围内评选 40 岁以下学者一名)。发表论文 100 余篇, Google Scholar 总引用 9000 余次。获得三次国际竞赛冠军以及四项最佳论文及提名奖。入选 IEEE Senior Member、国际期刊 IEEE Transactions on Information Forensics & Security (T-IFS, CCF:A 类) 副编, 中国图象图形学学会青托俱乐部副主席, 中国科学院青年创新促进会、北京科协青年人才托举工程、百度学术全球华人 AI 青年学者榜单 (全球 25 人), 受到腾讯犀牛鸟基金支持。获 2021 中国电子学会科技进步二等奖、中国图象图形学学会优秀博士论文提名奖。研发的三维人脸及人体动捕系统在阿里数字人生、淘宝直播、天猫数字藏品等 APP 上线; 研发的大规模人脸布控系统已在河南省公安厅、广州市公安局、无锡市公安局、福州市铁路公安处等单位应用; 研发的人脸身份认证服务平台应用于人民银行、太平人寿、证通股份等金融单位。

彭勃(bo.peng@nlpr.ia.ac.cn)



彭勃，博士，副研究员，2018 年至今在中科院自动化所工作。在图像内容取证与安全、深度伪造对抗攻防等领域有着深入研究经验。主持科技部重点研发课题、国家自然科学基金面上项目、中国科协青年托举项目、北京市自然科学基金面上项目等共 10 项。在 IEEE TIFS、TCSVT、CVPR、ACM MM 等重要国际期刊和会议上发表论文 30 余篇，获发明专利授权 17 项，软件著作权授权 5 项，曾获中国发明协会发明创业奖创新奖 一等奖、中国图象图形学学会技术发明奖二等奖等奖项和荣誉。目前是中国图象图形学会数字媒体取证与安全专业委员会委员、北京图象图形学会青工委委员、IEEE Beijing Biometrics Council Chapter Secretary。

李琦 (qli@nlpr.ia.ac.cn)



李琦，博士，副研究员，中国图象图形学会视觉大数据专委会委员，中国人工智能学会模式识别专委会委员。2016 年至今在中国科学院自动化研究所工作。主要研究方向包括计算机视觉、人脸预处理和识别等，主持国家重点研发计划青年科学家项目 1 项，主持国家自然科学基金面上基金和青年基金各 1 项，作为实际负责人承担国家重点研发计划课题、国家自然科学基金、北京自然科学基金等多项科研项目，参与的人脸识别系统成功应用于银行和铁路系统等。在国际主流期刊和会议发表 50 多篇文章，其中含 CCF-A 类文章 20 篇，以通讯作者发表的文章获评 CVPR 最佳论文提名，并被 TPAMI 邀约稿件，申请多项专利，现任 TPAMI, CVPR, NeurIPS, ICML 等审稿人。

赵健 (zhaoj90@chinatelecom.cn)



赵健，中国电信人工智能研究院多媒体认知学习实验室 (EVOL Lab) 负责人、青年科学家，西北工业大学光电与智能研究院研究员，博士毕业于新加坡国立大学，研究兴趣包括多媒体分析、临地安防、具身智能。围绕无约束视觉感知理解共发表 CCF-A 类论文 32 篇，以第一/通讯作者在 T-PAMI、CVPR 等国际权威期刊和会议上发表论文 31 篇，含一作 T-PAMI×2 (IF: 24.314)、IJCV×3 (IF: 13.369)，第一发明人授权国家发明专利 5 项。相关技术成果在百度、蚂蚁金服、奇虎 360 等 6 个科技行业领军企业得到应用，产生了显著效益。曾入选中国科协及北京市科协“青年人才托举工程”，主持国自然青年科学基金项目 6 项。曾获吴文俊人工智能优秀青年奖 (2023)、吴文俊人工智能自然科学奖一等奖 (2/5, 2022)、新加坡模式识别与机器智能协会 (PREMIA) Lee Hwee Kuan 奖、ACM Multimedia 唯一最佳学生论文奖 (一作, 1/208, CCF-A 类会议, 2018)，7 次在国际重要科技赛事中夺冠。担任北京图象图形学学会理事，国际知名期刊《Artificial Intelligence Advances》、《IET Computer Vision》编委，《Pattern Recognition Letters》、《Electronics》特刊客座编辑，VALSE 资深领域主席，ACM Multimedia 2021 分论坛主席，CICAI 2022/2023 领域主席，CCBR 2024 论坛主席，中国人工智能学会/中国图象图形学学会高级会员，“挑战杯”大学生科技作品竞赛评委，

中国人工智能大赛专家委员会等。

Structure-from-Motion: 大规模场景三维重建的基石
(组织者/主持人/讲者: 申抒含 中国科学院自动化研究所; 崔兆鹏 浙江大学;
崔海楠 中国科学院自动化研究所; 高翔 中国科学院自动化研究所)
(10月18日 15:30—17:15, 地点: 报告厅)

讲习班组织者/主持人/讲者简介

申抒含 (背景知识: 三维重建与 SfM 基础)



申抒含, 中国科学院自动化研究所研究员、博士生导师, 中国科学院大学人工智能学院岗位教授, 中国科学院工业视觉智能装备技术工程实验室副主任。研究领域为三维计算机视觉理论与应用, 包括大规模场景三维重建、智能机器人三维环境感知、场景三维语义理解等。在计算机视觉、摄影测量、机器人等领域国际期刊和国际会议, 如 IEEE Trans、IJCV、CVPR、ICCV、ICRA、IROS 等发表论文 90 余篇。作为项目负责人主持国家自然科学基金联合重点项目、北京市自然科学基金联合重点项目、以及各类企业研发项目 20 余项。入选中科院青年创新促进会、中科院自动化所特聘青年骨干、中科院特聘研究岗位, 曾获 2016 年 ACM 北京新星奖、2018 年中国图像图形学会科学技术二等奖、2023 年中国测绘学会科学技术一等奖、2023 年中国自动化学会自然科学一等奖等。

崔海楠 (解析求解: 经典 SfM 框架)



崔海楠, 中国科学院自动化研究所副研究员。2016 年博士毕业于中国科学院自动化研究所, 2011 年本科毕业于华南理工大学计算机科学与工程学院。主要从事三维计算机视觉理论和应用的研究, 尤其专注于基于图像的大规模场景三维建模研究, 已在 IEEE TIP、ISPRS P&RS、CVPR、AAAI、IROS 等国际期刊和会议发表论文 30 余篇。主要工作已被三维建模主流代码库 OpenMVG 和 TheiaSfM 实现和开源。主持国家自然科学基金青年项目/面上项目/联合重点项目课题、国家重点研发计划课题等, 同时得到华为、商汤、滴滴等国内企业长期资助, 获得 2020 年 CCF-滴滴盖亚青年学者科研基金杰出项目称号。

高翔 (解析求解: 混合 SfM 框架)



高翔, 中国科学院自动化研究所副研究员。2019 年于中国科学院自动化研究所获得博士学位, 2012 年、2015 年于中国海洋大学工程学院分别获得学士、硕士学位。研究方向为三维计算机视觉, 基于图像的大规模场景三维重建。在相关领域国际期刊 (IJCV、IEEE T-IP/T-NNLS/T-CYB/T-ITS/T-CSVT、ISPRS P&RS 等) 和国际会议 (CVPR、ECCV 等) 上发表论文 30 余篇, 主持国家自然科学基金青年/面上项目、中科院先导专项课题等国家级和省部级科研项目 7 项。所开发的基于图像的大规模场景三维重建系统已在城市高精地图构建、古建三维数字化保护、三维数字城市等多个领域得到了实际应用。

崔兆鹏（学习求解：隐式 SfM 框架）



崔兆鹏，浙江大学计算机科学与技术学院“百人计划”研究员、博士生导师，国家级青年人才计划入选者。2017年在加拿大西蒙弗雷泽大学获得博士学位。2017年至2020年在瑞士苏黎世联邦理工学院计算机视觉和几何实验室任高级研究员。研究方向为三维计算机视觉，主要从事基于视觉信息的三维感知和理解，具体包括三维重建、多视角立体几何、三维场景理解、同步定位与地图构建、三维内容生成等。近年来在计算机视觉、机器人、计算机图形学、机器学习等领域的顶级期刊和会议上发表论文50余篇，曾获ICRA 2020机器视觉最佳论文提名、IROS 2021安全、安保和救援机器人最佳论文提名、3DV 2024最佳论文荣誉提名。

口头报告一览表

场次	时间	编号	题目	会场
第一场	11:20-11:35	1	Efficient Fine-tuning Strategies for CLIP in Few-shot Scenarios via Supervised Contrastive Learning	丝路印象厅
	11:35-11:50	2	Two Semantic Information Extension Enhancement Methods For Zero-Shot Learning.	丝路印象厅
	11:50-12:05	3	Learning Fine-grained and Semantically Aware Mamba Representations for Tampered Text Detection in Images	丝路印象厅
	12:05-12:20	4	Multi-modality Correlation Learning Network for Pediatric Ventricular Septal Defects Identification	丝路印象厅
第二场	16:00-16:15	5	ADAL-GCN: Action Description Aided Learning Graph Convolution Network for Early Action Prediction	多功能厅
			H2LMER: A Cross Frame-Rate Representation Alignment Framework for Micro-Expression Recognition	学术厅
			Meta-Learning Based Knowledge Distillation for Domain Adaptive Nighttime Segmentation	塔里木厅
			Masked Visual Pre-training for RGB-D and RGB-T Salient Object Detection	报告厅
	16:15-16:30	6	Skeleton-Language Pre-training to Collaborate with Self-Supervised Human Action Recognition	多功能厅
			3DFaceMAE: Pre-training of Masked Autoencoder using Patch-based Random Masking Reconstruction and Super-resolution for 3D Face Recognition	学术厅
			Semantics Guided Disentangled GAN for Chest X-ray Image Rib Segmentation	塔里木厅
			Infrared Small Target Detection via Edge Refinement and Joint Attention Enhancement	报告厅
	16:30-16:45	7	Privacy-preserving Action Recognition: A Survey	多功能厅
			MFH: Marrying Frequency Domain with Handwritten Mathematical Expression Recognition	学术厅
			PottsNN: A Variational Neural Network Based on Potts Model for Image Segmentation	塔里木厅
			Multi-Branch Auxiliary Fusion YOLO with Re-parameterization Heterogeneous Convolutional for accurate object detection	报告厅
	16:45-17:00	8	Multi-scale Spatial and Temporal Feature Aggregation Graph Convolutional Network for Skeleton-Based Action Recognition	多功能厅
			Exploring Out-of-distribution Scene Text Recognition for Driving Scenes with Hybrid Test-time Adaptation	学术厅
			PRM: A Pixel-Region-Matching Approach for Fast Video Object Segmentation	塔里木厅
			Completing Saliency from Details	报告厅
17:00-17:15	9	Foreground-Background Partitioning and Feature Fusion for Weakly Supervised Fine-grained Image Recognition	多功能厅	
		Quat-DGNet: Enhancing 3D Dense Captioning with Quaternion-Based Spatial Offsets and Dynamic Neighborhood Graphs	学术厅	
		RT-VIS: Real-time Video Instance Segmentation with Light-weight Decoupled Framework	塔里木厅	
		Detect Text Forgery with Non-Forged Image Features: A Framework for Detection and Grounding of Image-Text Manipulation	报告厅	
第三场	11:20-11:35	10	A Novel Method for Autism Identification based on Multi-Atlas Features Fusion and Graph Neural Network	丝路印象厅
	11:35-11:50	11	From Point to Surface: Realistic and Perceptually-Plausible Hazy Image Generation with Glow-Diffusion	丝路印象厅
	11:50-12:05	12	Reducing Memory Footprint in Deep Network Training by Gradient Space Reutilization	丝路印象厅
	12:05-12:20	13	Dynamic Subframe Splitting and Spatio-Temporal Motion Entangled Sparse Attention for RGB-E Tracking	丝路印象厅

场次	时间	编号	题目	会场
第四场	16:00-16:15	14	DFANet: A Dual-stream Deep Feature Aware Network for Multi-focus Image Fusion	多功能厅
			MTFusion: Reconstructing Any 3D Object from Single Image Using Multi-Word Textual Inversion	报告厅
			Unleashing the Class-Incremental Learning Potential of Foundation Models by Virtual Feature Generation and Replay	塔里木厅
			Generative Steganography Based on Dual-Branch Flow	学术厅
	16:15-16:30	15	GAN-Diffusion Relay Model: Advancing Semantic Image Synthesis	多功能厅
			Disparity Refinement Based on Cross-Modal Feature Fusion and Global Hourglass Aggregation for Robust Stereo Matching	报告厅
			SLRL: Structured Latent Representation Learning for Multi-view Clustering	塔里木厅
			SACTGAN-EE imbalanced data processing method for credit default prediction	学术厅
	16:30-16:45	16	Multiview Light Field Angular Super-Resolution based on View Alignment and Frequency Attention	多功能厅
			M3Pose: Multi-person 3D Pose Estimation Using Sparse Millimeter-wave Radar Point Clouds	报告厅
			SPLICEGNN: SPLIt and ConnEct Tracklets in a unified Graph Neural Network	塔里木厅
			Misclassification Detection via Counterexample Learning for Trustworthy Cervical Cancer Screening	学术厅
	16:45-17:00	17	SFormer: Adaptive Sparse and Frequency-Guided Transformer Network for Single Image Derain	多功能厅
			Animatable Human Rendering from Monocular Video via Pose-Independent Deformation	报告厅
			Few-Shot Class-Incremental Learning via Cross-Modal Alignment with Feature Replay	塔里木厅
			DBMF-Net: A Dual-Branch Multimodal Fusion Network for Multi-Label Sewer Defect Classification	学术厅
	17:00-17:15	18	Hyperspectral Image Super-resolution Based on Dual-domain Gated Attention Network	多功能厅
			Discriminative-guided Diffusion-based Self-supervised Monocular Depth Estimation	报告厅
			Self-Quantization with Adaptive Codebooks for Unsupervised Image Retrieval	塔里木厅
			Feature Exchange and Distribution-based Mining Land Detection Method by Multispectral Imagery	学术厅

墙报论文一览表

海报 1		
序号	类别	论文题目
1	Medical Image Processing and Analysis	A Fine-grained Recurrent Network for Image Segmentation via Vector Field Guided Refinement
2	Medical Image Processing and Analysis	Semi-supervised Medical Image Segmentation with Strong/Weak Task-aware Consistency
3	Medical Image Processing and Analysis	Steerable Pyramid Transform Enables Robust Left Ventricle Quantification
4	Medical Image Processing and Analysis	MedPrompt: Cross-Modal Prompting for Multi-Task Medical Image Translation
5	Medical Image Processing and Analysis	Enhancing Hippocampus Segmentation: SwinUNETR Model Optimization with CPS
6	Medical Image Processing and Analysis	Uncertainty-inspired Credible Pseudo-Labeling in Semi-Supervised Medical Image Segmentation
7	Medical Image Processing and Analysis	MFPNet: Mixed Feature Perception Network for Automated Skin Lesion Segmentation
8	Medical Image Processing and Analysis	LD-BSAM: Combined Latent Diffusion with Bounding SAM for HIFU target region segmentation
9	Medical Image Processing and Analysis	Hierarchical Decoder with Parallel Transformer and CNN for Medical Image Segmentation
10	Medical Image Processing and Analysis	CLASS-AWARE CROSS PSEUDO SUPERVISION FRAMEWORK FOR SEMI-SUPERVISED MULTI-ORGAN SEGMENTATION IN ABDOMINAL CT SCANS
11	Medical Image Processing and Analysis	APAN: Anti-curriculum Pseudo-labelling and Adversarial Noises Training for Semi-supervised Medical Image Classification
12	Medical Image Processing and Analysis	Multi-Modal Learning for Predicting the Progression of Transarterial Chemoembolization Therapy in Hepatocellular Carcinoma
13	Medical Image Processing and Analysis	Growing with the help of multiple teachers: lightweight and noise-resistant student model for medical image classification
14	Medical Image Processing and Analysis	DRA-CN: A novel Dual-Resolution Attention Capsule Network for Histopathology Image Classification
15	Medical Image Processing and Analysis	A Mask Guided Network for Self-Supervised Low-Dose CT Imaging
16	Medical Image Processing and Analysis	Dental Diagnosis from X-Ray Panoramic Radiography Images: A Dataset and A Hybrid Framework
17	Medical Image Processing and Analysis	Edge-Guided Bidirectional-Attention Residual Network for Polyp Segmentation

18	Medical Image Processing and Analysis	From Coarse to Fine: A Novel Colon Polyp Segmentation Method Like Human Observation
19	Medical Image Processing and Analysis	Pseudo-Prompt Generating in Pre-trained Vision-Language Models for Multi-Label Medical Image Classification
20	Medical Image Processing and Analysis	Multi-Perspective Text-Guided Multimodal Fusion Network for Brain Tumor Segmentation
21	Medical Image Processing and Analysis	Continual Learning for Fundus Image Segmentation
22	Medical Image Processing and Analysis	Embedded Deep Learning Based CT Images for Rifampicin Resistant Tuberculosis Diagnosis
23	Medical Image Processing and Analysis	Combining Segment Anything Model with Domain-Specific Knowledge for Semi-Supervised Learning in Medical Image Segmentation
24	Medical Image Processing and Analysis	Meply: A Large-scale Dataset and Baseline Evaluations for Metastatic Perirectal Lymph Node Segmentation
25	Medical Image Processing and Analysis	Swin-HAUnet: A Swin-Hierarchical Attention Unet For Enhanced Medical Image Segmentation
26	Medical Image Processing and Analysis	ODC-SA Net: Orthogonal Direction Enhancement and Scale Aware Network for Polyp Segmentation
27	Medical Image Processing and Analysis	Two-Stage Multi-Scale Feature Fusion for Small Medical Object Segmentation
28	Medical Image Processing and Analysis	A Two-Stage Automatic Collateral Scoring Framework Based on Brain Vessel Segmentation
29	Medical Image Processing and Analysis	SPARK: Cross-Guided Knowledge Distillation with Spatial Position Augmentation for Medical Image Segmentation
30	Medical Image Processing and Analysis	VATBoost-Net: Integrating Enhanced Feature Perturbation and Detail Enhancement for Medical Image Segmentation
31	Medical Image Processing and Analysis	DTIL-Net: Dual-Task Interactive Learning Network for Automated Grading of Diabetic Retinopathy and Macular Edema
32	Medical Image Processing and Analysis	DeformSegNet: Segmentation Network Fused with Deformation Field for Pancreatic CT Scans
33	Medical Image Processing and Analysis	InsSegLN: A Novel 3D Instance Segmentation Method for Mediastinal Lymph Node
34	Medical Image Processing and Analysis	RRANet: A Reverse Region-Aware Network with Edge Difference for Accurate Breast Tumor Segmentation in Ultrasound Images
35	Medical Image Processing and Analysis	Learning Frequency and Structure in UDA for Medical Object Detection
36	Medical Image Processing and Analysis	Skin Lesion Segmentation Method Based On Global Pixel Weighted Focal Loss

37	Medical Image Processing and Analysis	Competing Dual-Network with Pseudo-Supervision Rectification for Semi-Supervised Medical Image Segmentation
38	Medical Image Processing and Analysis	Dual-Branch Perturbation and Conflict-Based Scribble-Supervised Meibomian Gland Segmentation
39	Medical Image Processing and Analysis	Anchored Supervised Contrastive Learning for Long-Tailed Medical Image Regression
40	Medical Image Processing and Analysis	Dynamic Feature Fusion Based on Consistency and Complementarity of Brain Atlases
41	Medical Image Processing and Analysis	FUF-TransUNet: a transformer-based U-Net with fully utilize of features for liver and liver-tumor segmentation in CT images
42	Medical Image Processing and Analysis	Dual-View Dual-Boundary Dual U-Nets for Multiscale Segmentation of Oral CBCT Images
43	Medical Image Processing and Analysis	A Novel Diffusion Model with Wavelet Transform for Optic Disc and Cup Segmentation in Fundus Images
44	Medical Image Processing and Analysis	STCTb: A Spatio-Temporal Collaborative Transformer Block for Brain Diseases Classification using fMRI Time Series.
45	Medical Image Processing and Analysis	A Generalized Contrast-adjustment Guided Growth Method for Medical Image Segmentation
46	Medical Image Processing and Analysis	MDNet: Morphology-Driven Weakly Supervised Polyp Detection
47	Medical Image Processing and Analysis	MMR-Sleep: A Multi-Channel and Multi-Receptive Field Sleep Stage recognition Model
48	Medical Image Processing and Analysis	CPNet: Cross Prototype Network for Few-shot Medical Image Segmentation
49	Medical Image Processing and Analysis	SBC-UNet: A Network Based on Improved Hourglass Attention Mechanism and U-Net for Medical Image Segmentation
50	Medical Image Processing and Analysis	Bridge the gap of semantic context: A Boundary-guided Context Fusion UNet for Medical Image Segmentation
51	Medical Image Processing and Analysis	Bilinear Fine-grained Classification of Ultrasound Images Integrated with Interpretable Radiomics
52	Medical Image Processing and Analysis	GCNet: Global context-guided uncertainty boundary for polyp segmentation
53	Medical Image Processing and Analysis	Comprehensive Transformer Integration Network (CTIN): Advancing Endoscopic Disease Segmentation with Hybrid Transformer Architecture
54	Medical Image Processing and Analysis	IPM: An Intelligent Component for 3D Brain Tumor Segmentation Integrating Semantic Extractor and Pixel Refiner
55	Medical Image Processing and Analysis	Edge-Net: A Self-supervised Medical Image Segmentation Model Based on Edge Attention

56	Medical Image Processing and Analysis	Fundus image disease diagnosis and quality assessment based on dual-task collaborative optimization
57	Medical Image Processing and Analysis	MFIS-net: A Deep Learning Framework for Left Atrial Segmentation
58	Medical Image Processing and Analysis	Semi-Supervised Gland Segmentation via Label Purification and Reliable Pixel Learning
59	Multi-Modal Information Processing	A Multi-modal Framework with Contrastive Learning and Sequential Encoding for Enhanced Sleep Stage Detection
60	Multi-Modal Information Processing	Charting the Uncharted: Building and Analyzing a Multifaceted Chart Question Answering Dataset for Complex Logical Reasoning Process
61	Multi-Modal Information Processing	Time-Frequency Mutual Learning for Moment Retrieval and Highlight Detection
62	Multi-Modal Information Processing	Cascade Coarse-to-Fine Point-Query Transformer for RGB-T Crowd Counting
63	Multi-Modal Information Processing	Perceptual Image Compression with Text-Guided Multi-Level Fusion
64	Multi-Modal Information Processing	Evaluating Attribute Comprehension in Large Vision-language Models
65	Multi-Modal Information Processing	Efficient Multi-modal Human-centric Contrastive Pre-training with A Pseudo Body-structured Prior
66	Multi-Modal Information Processing	A3R: Vision Language Pre-training by Attentive Alignment and Attentive Reconstruction
67	Multi-Modal Information Processing	Mixture-of-Hand-Experts: Repainting the Deformed Hand Images Generated by Diffusion Models
68	Multi-Modal Information Processing	ConD2: Contrastive Decomposition Distilling for Multimodal Sentiment Analysis
69	Multi-Modal Information Processing	Multi-layer Tuning CLIP for Few-Shot Image Classification
70	Multi-Modal Information Processing	DIM: Dynamic Integration of Multimodal Entity Linking with Large Language Model
71	Multi-Modal Information Processing	Text-Dominant Interactive Attention for Cross-Modal Sentiment Analysis.
72	Multi-Modal Information Processing	Dual Context Perception Transformer for Referring Image Segmentation
73	Multi-Modal Information Processing	ELEMO: Elements Focused Emotion Recognition for Sticker Images
74	Multi-Modal Information Processing	Cross-Modal Dual Matching and Comparison for Text-to-Image Person Re-identification
75	Multi-Modal Information Processing	Low-resource Machine Translation with Different Granularity Image Features
76	Multi-Modal Information Processing	ST-SBV: Spatial-Temporal Self-Blended Videos for Deepfake Detection

77	Multi-Modal Information Processing	Learning a Robust Synthetic Modality with Dual-Level Alignment for Visible-Infrared Person Re-identification
78	Multi-Modal Information Processing	Deep Noisy Multi-Label Learning for Robust Cross-Modal Retrieval
79	Multi-Modal Information Processing	Uncertainty-Aware with Negative Samples for Video-Text Retrieval
80	Multi-Modal Information Processing	Multi-Modal Knowledge-enhanced Fine-Grained Image Classification
81	Multi-Modal Information Processing	Bridging Modality Gap for Visual Grounding with Effective Cross-modal Distillation
82	Multi-Modal Information Processing	EGSRNet: Emotion-label Guiding and Similarity Reasoning Network for Multimodal Sentiment Analysis
83	Multi-Modal Information Processing	VL-MPFT: Multitask Parameter-Efficient Fine-Tuning for Visual-Language Pre-trained Models via Task-adaptive Masking
84	Multi-Modal Information Processing	A Multimodal Fake News Detection Model Leveraging Image Frequency and Spatial Domain Analysis with Deep Dynamic Trade-off Fusion
85	Multi-Modal Information Processing	Efficiency-Aware Fine-grained Vision-Language Retrieval via a Global-Contextual Autoencoder
86	Multi-Modal Information Processing	Towards Making the Most of Knowledge across Languages for Multimodal Cross-Lingual Summarization
87	Multi-Modal Information Processing	Adapting Vision-Language Models to Open Classes via Test-Time Prompt Tuning
88	Multi-Modal Information Processing	Multimodal Feature Hierarchical Alignment for Text-Based Person Re-identification
89	Multi-Modal Information Processing	Mitigating Hallucination in Visual-Language Models via Re-Balancing Contrastive Decoding
90	Multi-Modal Information Processing	Multimodal medical image registration using optimized phase consistency within joint Frequency-Space domain
91	Multi-Modal Information Processing	Robust Contrastive Learning against Audio-Visual Noisy Correspondence
92	Multi-Modal Information Processing	Enhancing Cross-Modal Alignment in Multimodal Sentiment Analysis via Prompt Learning
93	Multi-Modal Information Processing	Efficient Language-driven Action Localization by Feature Aggregation and Prediction Adjustment
94	Multi-Modal Information Processing	Greedy Fusion Oriented Representations for Multimodal Sentiment Analysis
95	Multi-Modal Information Processing	Counterfactual Multimodal Fact-Checking Method based on Causal Intervention
96	Multi-Modal Information Processing	Rethinking the Necessity of Learnable Modal Alignment for Medical Image Fusion

97	Multi-Modal Information Processing	Taming Diffusion for Fashion Clothing Generation with Versatile Condition
98	Video Analysis and Understanding	Behavior Capture Based Explainable Engagement Recognition
99	Video Analysis and Understanding	Utilizing Text-video Relationships: A Text-driven Multi-modal Fusion Framework for Moment Retrieval and Highlight Detection
100	Video Analysis and Understanding	LLMAAction: Adapting Large Language Model for Long-Term Action Anticipation
101	Video Analysis and Understanding	Dual-scale Temporal Dependency Learning for Unsupervised Video Anomaly Detection
102	Video Analysis and Understanding	LMS-VDR: Integrating Landmarks into Multi-Scale Hybrid Net for Video-based Depression Recognition
103	Video Analysis and Understanding	Motion Trajectory Reconstruction Based on Feature Matching and Gradient Graph Laplacian Regularizer
104	Video Analysis and Understanding	OC-SAN: Unsupervised Deepfake Detection for Specific Individual Protection Based on Deep One-class Classification
105	Video Analysis and Understanding	Focus on Subtle Actions: Semantic and Saliency Knowledge Co-Propagation Method for Weakly-Supervised Temporal Action Localization
106	Video Analysis and Understanding	Flow-Audio-Synth:A Video-to-Audio Model which Captures Dynamic Features
107	Video Analysis and Understanding	SCAMS: Semantic Category-Aware Multi-Scale Network for Video Quality Assessment
108	Video Analysis and Understanding	Cross-temporal Fusion Memory Network for Traffic Accident Detection
109	Video Analysis and Understanding	Video Frame Interpolation for Large Motion with Generative Prior
110	Video Analysis and Understanding	MS-DETR: Exploiting Modality Synergy for Moment Retrieval and Highlight Detection
111	Video Analysis and Understanding	PosCap: Boosting Video Captioning with Part-of-Speech Guidance
112	Video Analysis and Understanding	Element-Centered Multi-Granularity Network for Dense Video Captioning
113	Video Analysis and Understanding	AI-Generated Video Detection via Spatial-Temporal Anomaly Learning
114	Video Analysis and Understanding	Dynamic Temporal Shift Feature Enhancement for Few-Shot Action Recognition
115	Multimedia Analysis and Reasoning	StyleFusion TTS: Multimodal Style-control and Enhanced Feature Fusion for Zero-shot Text-to-speech Synthesis
116	Multimedia Analysis and Reasoning	Robust Document Presentation Attack Detection via Diffusion Models and Knowledge Distillation
117	Multimedia Analysis and Reasoning	Towards the transferable reversible adversarial example via distribution-relevant attack

118	Multimedia Analysis and Reasoning	Fine-grained Feature Assisted Cross-modal Image-text Retrieval
119	Multimedia Analysis and Reasoning	Uncertainty-aware Gradient Modulation and Feature Masking for Multimodal Sentiment Analysis
120	Multimedia Analysis and Reasoning	Local and Global Features Interactive Fusion Network for Macro- and Micro-Expression Spotting in Long Videos
121	Multimedia Analysis and Reasoning	OmniStyleGAN for Style-Guided Image-to-Image Translation
122	Fundamental Theory of Computer Vision	Miss-CAM: Visual interpretation algorithm for convolutional neural networks using missingness masks
123	Fundamental Theory of Computer Vision	GSE-Ships: Ship Detection Using Optimized Lightweight Networks and Attention Mechanisms
124	Fundamental Theory of Computer Vision	Balancing Complementarity and Consistency via Delayed Activation in Incomplete Multi-view Clustering
125	Fundamental Theory of Computer Vision	Small Target Defects Detection of Aluminum Plates Sur-face using an MSN-YOLOv5 model
126	Performance Evaluation and Benchmarks	Benchmarking Multi-Scene Fire and Smoke Detection
127	Performance Evaluation and Benchmarks	Performance Evaluation of Anomaly Detection with a new Battery Surface Anomaly Dataset
128	Performance Evaluation and Benchmarks	Fine-grained Metrics for Point Cloud Semantic Segmentation
129	Performance Evaluation and Benchmarks	114Xray: A Large-scale X-ray Security Detection Benchmark and Aware Enhance Network for Real-World Prohibited Item Inspection in Baggage
130	Structural Pattern Recognition	Enhancing Semi-Dense Feature Matching through Probabilistic Modeling of Cascaded Supervision and Consistency
131	Structural Pattern Recognition	Concentrating Estimation Attention: Human Prior Constrained Methods for Robust Classification

海报 2		
序号	类别	论文题目
1	Object Detection Tracking and Identification	Mamba-FETrack: Frame-Event Tracking via State Space Model
2	Object Detection Tracking and Identification	Stay Open: Calibrating Weights Continuously for Detecting Out-of-Distribution Objects On the Fly
3	Object Detection Tracking and Identification	MPE: A Fine-grained Multi-Path Feature Enhancer in MOT
4	Object Detection Tracking and Identification	CIMTD: Class Incremental Multi-Teacher Knowledge Distillation for Fractal Object Detection
5	Object Detection Tracking and Identification	A Stochastic Model for video object tracking

6	Object Detection Tracking and Identification	Class-Agnostic Detection of Unknown Objects From Foreground Improves Robust Open World Object Detection
7	Object Detection Tracking and Identification	ML-SCODNet: Multitask Learning for Scene Classification and Object Detection Network from Remote Sensing Images
8	Object Detection Tracking and Identification	Cross-Domain Attention Alignment for Domain Adaptive Person Re-ID
9	Object Detection Tracking and Identification	EfficientMatting: Bilateral Matting Network for Real-time Human Matting
10	Object Detection Tracking and Identification	Semi-Supervised Camouflaged Object Detection: Multi Information Fusion Combined with Adaptive Receptive Field Selection Network
11	Object Detection Tracking and Identification	Revisiting Network Perturbation for Semi-Supervised Semantic Segmentation
12	Object Detection Tracking and Identification	TL-ReID: Tight-Loose Pairwise Loss for Object Re-Identification
13	Object Detection Tracking and Identification	LS-YOLO: A Lightweight Selective Enhanced YOLOv8 Algorithm for UAV Aerial Photography
14	Object Detection Tracking and Identification	Camouflaged Object Detection via Scale-Feature Attention and Type-Feature Attention
15	Object Detection Tracking and Identification	GOP: A Group Object Perception Framework for Optical Remote Sensing
16	Object Detection Tracking and Identification	Mask-Guided Clothes-irrelevant and Background-irrelevant Network with Knowledge Propagation for Cloth-Changing Person Re-Identification
17	Object Detection Tracking and Identification	Region Aware Transformer with Intra-Class Compact for Unsupervised Aerial Person Re-identification
18	Object Detection Tracking and Identification	Camouflaged Object Detection based on Feature Aggregation and Global Semantic Learning
19	Object Detection Tracking and Identification	Small Target Detector based on Adaptive Re-parameterized Spatial Feature Fusion Mechanism
20	Object Detection Tracking and Identification	Dual Constraint Parallel Multi-scale Attention Network for Insulator Detection in Foggy Scene
21	Object Detection Tracking and Identification	Lightweight Defog Detection for Autonomous Vehicles: Balancing Clarity, Efficiency, and Accuracy
22	Object Detection Tracking and Identification	EI-YOLO: Efficiently Improved YOLO on Detection of Prohibited Items During Security inspections
23	Object Detection Tracking and Identification	TSTrack: A Robust Object Tracking Framework Integrated Temporal and Spatial Features
24	Object Detection Tracking and Identification	A Faster Fire Detection Network with Global Information Awareness
25	Object Detection Tracking and Identification	More Efficient Encoder: Boosting Transformer-Based Multi-Object Tracking Performance Through YOLOX

26	Object Detection Tracking and Identification	PLRUT: Pseudo Label and Re-detection boosted Unsupervised Tracking of Unmanned Aerial Vehicle Objects
27	Object Detection Tracking and Identification	End-to-end High-quality Transformer Object Detection Model Applied to Human Head Detection
28	Object Detection Tracking and Identification	Local Point Matching for Collaborative Image Registration and RGBT Anti-UAV Tracking
29	Object Detection Tracking and Identification	ORU-YOLO: A UAV Image Detection Model Optimized for Resource Utilization
30	Object Detection Tracking and Identification	Memoryless Multimodal Anomaly Detection via Student-Teacher Network and Signed Distance Learning
31	Object Detection Tracking and Identification	MRFNet: A Multi-Receptive-Field Fusion Network for Multi-Food Recognition
32	Object Detection Tracking and Identification	Image-Centered Pseudo Label Generation for Weakly Supervised Text-based Person Re-Identification
33	Object Detection Tracking and Identification	SelfLoc: High Quality Unsupervised Object Localization with Self-Prompt SAM
34	Object Detection Tracking and Identification	A Global Re-detection Method Based on Feature Interaction Siamese Network
35	Object Detection Tracking and Identification	Key Object Detection: Unifying Salient and Camouflaged Object Detection into One Task
36	Object Detection Tracking and Identification	MOFTrack: Multi-Object Formation Tracking in Remote Sensing Videos
37	Object Detection Tracking and Identification	SDNet: A Simple and Efficient Salient Object Detection Decoder with Only 60K Parameters
38	Object Detection Tracking and Identification	Scale-Adaptive Modulation Meet Compact Axial Transformer for Small Object Detection in UAV-Vision
39	Object Detection Tracking and Identification	Lightweight and Multi-Scale Adaptive Network for Infrared Small Target Detection
40	Object Detection Tracking and Identification	Multi-View Cross-Attention Network for Hyperspectral Object Tracking
41	Object Detection Tracking and Identification	CountMamba: Exploring Multi-directional Selective State-Space Models for Plant Counting
42	Object Detection Tracking and Identification	ECLNet: A Compact Encoder-Decoder Network for Efficient Camouflaged Object Detection
43	Object Detection Tracking and Identification	Few-Shot Object Detection via Disentangling Class-Related Factors in Feature Distribution
44	Object Detection Tracking and Identification	Multi-class token-guided end-to-end weakly supervised image semantic segmentation method
45	Object Detection Tracking and Identification	DIDNet: An End-to-End Directional Insulator Detection Network based on direction field
46	Object Detection Tracking and Identification	L2FIG-Tracker: l_2 -norm based Fusion with Illumination Guidance for RGB-D Object Tracking

47	Object Detection Tracking and Identification	CDAF3D: Cross-Dimensional Attention Fusion for Indoor 3D Object Detection
48	Object Detection Tracking and Identification	RETrack: Multi-Object Tracking by Associating Proposal Regions
49	Object Detection Tracking and Identification	PGNET: A Real-time efficient model for underwater object detection
50	Object Detection Tracking and Identification	A Temporal Recognition Framework for Multi-Sheep Behaviour Using ViTSORT and YOLOv8-MS
51	Object Detection Tracking and Identification	Tracking Transforming Objects: A Benchmark
52	Object Detection Tracking and Identification	Modality-Shared Prototypes for Enhanced Unsupervised Visible-Infrared Person Re-identification
53	Object Detection Tracking and Identification	Vehicle Re-identification with a Pose-aware Discriminative Part Learning Model
54	Object Detection Tracking and Identification	Dual-Teacher Network with SSIM based Reverse Distillation for Anomaly Detection
55	Object Detection Tracking and Identification	CFMVOR: Federated Multi-view 3D Object Recognition Based on Compressed Learning
56	Object Detection Tracking and Identification	Enhanced Anomaly Detection using Spatial-Alignment and Multi-scale Fusion
57	Object Detection Tracking and Identification	Confidence-Weighted Teacher: Semi-Supervised Object Detection Based on Confidence Correction
58	Biometric Recognition	MST-Gait: Application of Multi-Scale Temporal Modeling to Gait Recognition
59	Biometric Recognition	Identity-Preserving Animal Image Generation for Animal Individual Identification
60	Biometric Recognition	FIL-FLD: Few-shot Incremental Learning with EMD Metric for High Generalization Fingerprint Liveness Detection
61	Biometric Recognition	Text Based Unsupervised Domain Generalization Person Re-identification
62	Biometric Recognition	SF-Gait: Two-Stage Temporal Compression Network for Learning Gait Micro-Motions and Cycle Patterns
63	Biometric Recognition	Coarse-to-Fine Domain Adaptation for Cross-subject EEG Emotion Recognition with Contrastive Learning
64	Biometric Recognition	Face Anti-spoofing based on Multi-view Anomaly Detection
65	Biometric Recognition	Online Signature Verification Based on Recurrent Attentional Time-Delay Neural Networks
66	Biometric Recognition	Multimodal finger recognition based on feature fusion attention for fingerprints, finger-veins, and finger-knuckle-prints
67	Biometric Recognition	Hierarchical Discrepancy-aware Interaction Network for Face Forgery Detection

68	Biometric Recognition	AU-vMAE: Knowledge-Guide Action Units Detection via Video Masked Autoencoder
69	Biometric Recognition	Transformer-based Multimodal Spatial-Temporal Fusion for Gait Recognition
70	Biometric Recognition	Multi-level Distributional Discrepancy Enhancement for Cross Domain Face Forgery Detection
71	Biometric Recognition	Unsupervised person Re-ID based on nonlinear asymmetric metric learning
72	Biometric Recognition	FR-watermarking: A Fusion Framework for Face-Based Digital Watermarking
73	Document Analysis and Recognition	Leveraging Structure Knowledge and Deep Models for the Detection of Abnormal Handwritten Text
74	Document Analysis and Recognition	OCR-aware Scene Graph Generation via Multi-modal Object Representation Enhancement and Logical Bias Learning
75	Document Analysis and Recognition	Enhancing Transformer-based Table Structure Recognition for Long Tables
76	Document Analysis and Recognition	Show Exemplars and Tell Me What You See: In-context Learning with Frozen Large Language Models for TextVQA
77	Document Analysis and Recognition	MLR-NET: an arbitrary skew angle detection algorithm for complex layout document images
78	Document Analysis and Recognition	TextViTCNN: Enhancing Natural Scene Text Recognition with Hybrid Transformer and Convolutional Networks
79	Document Analysis and Recognition	Enhancing Visual Information Extraction with Large Language Models through Layout-aware Instruction Tuning
80	Document Analysis and Recognition	SFENet: Arbitrary Shapes Scene Text Detection with Semantic Feature Extractor
81	Document Analysis and Recognition	Improving Zero-Shot Image Captioning Efficiency with Metropolis-Hastings Sampling
82	Document Analysis and Recognition	Improving Text Classification Performance through Multimodal Representation
83	Document Analysis and Recognition	A Multi-feature Fusion Approach for Words Recognition of Ancient Mongolian Documents
84	Document Analysis and Recognition	TableRocket: An Efficient and Effective Framework for Table Reconstruction
85	Document Analysis and Recognition	Not All Texts Are the Same: Dynamically Querying Texts for Scene Text Detection
86	Document Analysis and Recognition	Multi-Modal Attention based on 2D Structured Sequence for Table Recognition
87	Character Recognition	Scene Text Recognition via k-NN Attention-based Decoder and Margin-based Softmax Loss
88	Character Recognition	Real-Time Text Detection with Multi-Level Feature Fusion and Pixel Clustering

89	Character Recognition	REFINED AND LOCALITY-ENHANCED FEATURE FOR HANDWRITTEN MATHEMATICAL EXPRESSION RECOGNITION
90	Character Recognition	Dual Feature Enhanced Scene Text Recognition Method for Low-Resource Uyghur
91	Character Recognition	Segmentation-free Todo Mongolian OCR and Its Public Dataset
92	Character Recognition	Hybrid Encoding Method for Scene Text Recognition in Low-Resource Uyghur
93	Character Recognition	ROBC: a Radical-Level Oracle Bone Character Dataset
94	Character Recognition	Integrated Recognition of Arbitrary-Oriented Multi-Line Billet Number
95	Character Recognition	Improving Scene Text Recognition with Counting Aware Contrastive Learning and Attention Alignment
96	Character Recognition	GridMask: An Efficient Scheme for Real Time Curved Scene Text Detection
97	Character Recognition	Tibetan Handwriting Recognition Method based on Structural Re-parameterization ViT and Vertical Attention
98	Face Recognition and Pose Recognition	MPM: A Unified 2D-3D Human Pose Representation via Masked Pose Modeling
99	Face Recognition and Pose Recognition	Joint Multi-Cue Learning for Emotion Recognition in Human-Computer Interaction
100	Face Recognition and Pose Recognition	Depth Decoupling for Bottom-Up Multi-Person 3D Pose Estimation
101	Face Recognition and Pose Recognition	TT-DF: A Large-Scale Diffusion-Based Dataset and Benchmark for Human Body Forgery Detection
102	Face Recognition and Pose Recognition	Walking is Matter: A Benchmark for Fine-Grained Gait Segmentation
103	Face Recognition and Pose Recognition	Spatial-Frequency Dual-stream Reconstruction for Deepfake Detection
104	Face Recognition and Pose Recognition	DeepSweep: Real-Time Multi-View 3D Pose Estimation via Cross-View Deep Matching and Plane Sweeping
105	Face Recognition and Pose Recognition	PoseVR:Structure-aware Hybrid Full-Body Pose Estimation in Virtual Reality
106	Face Recognition and Pose Recognition	Fusion Network Based on Motion Learning and Image Feature Representation for Micro-expression Recognition
107	Face Recognition and Pose Recognition	Depth-Aware Dual-stream Interactive Transformer Network for Facial Expression Recognition
108	Face Recognition and Pose Recognition	SCALE-Pose: Skeletal Correction and Language Knowledge-assisted for 3D Human Pose Estimation
109	Action Recognition	A Two-stream Hybrid CNN-Transformer Network for Skeleton-based Human Interaction Recognition

110	Action Recognition	Spatio-Temporal Contrastive Learning for Compositional Action Recognition
111	Action Recognition	Path-Guided Motion Prediction with Multi-View Scene Perception
112	Action Recognition	Attention-based Spatio-temporal modeling with 3D Convolutional Neural Networks for Dynamic Gesture Recognition
113	Action Recognition	MIT: Multi-cue Injected Transformer for Two-stage HOI Detection
114	Action Recognition	DIDA: Dynamic Individual-to-integrated Augmentation for Self-Supervised Skeleton-Based Action Recognition
115	Action Recognition	Improving Video Representation of Vision-Language Model with Decoupled Explicit Temporal Modeling
116	Action Recognition	KS-FuseNet: An efficient action recognition method based on keyframe selection and feature fusion
117	Action Recognition	Dynamic Skeleton Association Transformer for dyadic Interaction Action Recognition
118	Action Recognition	Species-Aware Guidance for Animal Action Recognition with Vision-Language Knowledge
119	Feature Extraction and Feature Selection	Credit-based Negative Sample Denoising in Contrastive Learning
120	Feature Extraction and Feature Selection	UDD: Dataset Distillation via Mining Underutilized Regions
121	Feature Extraction and Feature Selection	BertTab: Table Learning with Feature Descriptions and Context
122	Feature Extraction and Feature Selection	SemVG: Semantic Fused Feature Extraction Network for Visual Geo-localization under Urban Street Scenes
123	Feature Extraction and Feature Selection	Efficient Discriminative Feature Selection With Grouping Relative Comparison
124	Feature Extraction and Feature Selection	Block Cipher Algorithm Identification Based On CNN-Transformer Fusion Model
125	Feature Extraction and Feature Selection	DSTF: Dual-Stream Spatio-Temporal Fusion Network for Event-Based Data
126	Pattern Classification and Cluster Analysis	Adapt and Refine: A Few-Shot Class-Incremental Learner via Pre-trained Models
127	Pattern Classification and Cluster Analysis	Learning Fully Parametric Subspace Clustering
128	Pattern Classification and Cluster Analysis	A Comprehensive Exploration on Detecting Fake Images Generated by Stable Diffusion
129	Pattern Classification and Cluster Analysis	Adaptive Margin Global Classifier for Exemplar-Free Class-Incremental Learning
130	Pattern Classification and Cluster Analysis	FedHC: Learning Imbalanced Clusters via Federated Hierarchical Clustering

131	Pattern Classification and Cluster Analysis	Enhancing Time Series Classification with Explainable Time-frequency Features Representation
132	Pattern Classification and Cluster Analysis	Adaptive Unified Framework with Global Anchor Graph for Large-scale Multi-view Clustering

海报 3		
序号	类别	论文题目
1	Low-Level Vision and Image Processing	Focal Perception Transformer for Light Field Salient Object Detection
2	Low-Level Vision and Image Processing	A Fourier Transform Framework for Domain Adaptation
3	Low-Level Vision and Image Processing	Bidirectional Alternating Fusion Network for RGB-T Salient Object Detection
4	Low-Level Vision and Image Processing	SimpleFusion: A Simple Fusion Framework for Infrared and Visible Images
5	Low-Level Vision and Image Processing	A Cross-Consistency Strategy for Clearer Perception in Low-Light Haze
6	Low-Level Vision and Image Processing	Improve Segmentation robustness of Intracellular Structures in Fluorescence Microscopy Images
7	Low-Level Vision and Image Processing	RDSR:Reparameterized Lightweight Diffusion Model For Image Super-Resolution
8	Low-Level Vision and Image Processing	ACENet: Adaptive Context Enhancement Network for RGB-T Video Object Detection
9	Low-Level Vision and Image Processing	Fake-GPT: Detecting Fake Image via Large Language Model
10	Low-Level Vision and Image Processing	Towards Elastic Image Super-Resolution Network via Progressive Self-distillation
11	Low-Level Vision and Image Processing	PHANet: Progressive Hybrid Attention Network for Enhanced Video Deraining
12	Low-Level Vision and Image Processing	Frequency Adapter and Spatial Prompt Network for All-in-One Blind Image Restoration
13	Low-Level Vision and Image Processing	Ambient Illumination Disentangled Based Weakly-Supervised Image Restoration Using Adaptive Pixel Retention Factor
14	Low-Level Vision and Image Processing	F4SR: A Feed-Forward Regression Approach for Few-Shot Face Super-Resolution
15	Low-Level Vision and Image Processing	Thangka Mural Super-Resolution Based on Nimble Convolution and Overlapping Window Transformer
16	Low-Level Vision and Image Processing	EHAT:Enhanced Hybrid Attention Transformer for Remote Sensing Image Super-Resolution
17	Low-Level Vision and Image Processing	MIAFusion: Infrared and Visible Image Fusion via Multi-scale Spatial and Channel-Aware Interaction Attention
18	Low-Level Vision and Image Processing	Color Enhanced Network for Image Dehazing

19	Low-Level Vision and Image Processing	Advancing Real-World Burst Denoising: A New Benchmark and Dual-Branch Burst Denoising Network
20	Low-Level Vision and Image Processing	DIFNet: Dual-Domain Information Fusion Network for Image Denoising
21	Low-Level Vision and Image Processing	Simultaneous Snow Mask Prediction and Single Image Desnowing with a Bidirectional Attention Transformer Network
22	Low-Level Vision and Image Processing	DBIF: Dual-Branch Feature Extraction Network for Infrared and Visible Image Fusion
23	Low-Level Vision and Image Processing	Multi-dimensional Information Awareness Residual Network for Lightweight Image Super-Resolution
24	Low-Level Vision and Image Processing	Feature Pruning and Recovery Learning with Knowledge Distillation for Occluded Person Re-Identification
25	Low-Level Vision and Image Processing	A Redundancy-suppression based Event Sampling Method for Structured Representation
26	Low-Level Vision and Image Processing	BSDiff: Low-Light Image Enhancement Via Blueprint Separable Convolution and Wavelet-Diffusion Model
27	Low-Level Vision and Image Processing	SAM and Diffusion Based Adversarial Sample Generation for Image Quality Assessment
28	Low-Level Vision and Image Processing	CF-LAM: Coarse-to-Fine Locally Affine Matching for Viewpoint Transformations
29	Low-Level Vision and Image Processing	FormerUnify: Transformer-Based Unified Fusion for Efficient Image Matting
30	Low-Level Vision and Image Processing	Camouflage Object Segmentation with Multi-scale Feature Aggregation and Boundary Generation
31	Low-Level Vision and Image Processing	GCMLP: A Lightweight Network for Gamut Compression
32	Low-Level Vision and Image Processing	Low-Light Light-Field Image Enhancement with Geometry Consistency
33	Low-Level Vision and Image Processing	Attention and Boundary Induced Feature Refinement Network for Camouflaged Object Detection
34	Low-Level Vision and Image Processing	Towards Specular Highlight Removal Through Diffusion Model
35	Low-Level Vision and Image Processing	FIR: A plug-in Feature-to-Image Reconstruction Method for Feature Coding for Machines
36	Low-Level Vision and Image Processing	Focal Aggregation Transformer for Light Field Image Super-Resolution
37	Low-Level Vision and Image Processing	CoMoFusion: Fast and High-quality Fusion of Infrared and Visible Image with Consistency Model
38	Low-Level Vision and Image Processing	Patch Attacks on Vision Transformer via Skip Attention Gradients
39	Low-Level Vision and Image Processing	Multi-scale Progressive Reconstruction Network for High Dynamic Range Imaging

40	Low-Level Vision and Image Processing	Fine-Grained Adjustable Entropy Models for Rate-Complexity Jointly Adjustable Image Compression
41	Low-Level Vision and Image Processing	FPSNet: Focus-Perceptual-Semantic Full Flow Visual Redundancy Predicting for Camera Image
42	Low-Level Vision and Image Processing	Semantic-Aware Global and Local Fusion Model for Image Enhancement
43	Low-Level Vision and Image Processing	A Two-branch Fusion Network for Infrared and Visible Image Fusion
44	Low-Level Vision and Image Processing	Attention-Guided Residual Fourier Transformation Network for Single Image Deblurring
45	Low-Level Vision and Image Processing	I3En: A Multi-Level Iterative Low-Light Enhancement Network Based on Sketch Prior Guidance
46	Low-Level Vision and Image Processing	Efficient Stereo Matching Using Dynamic Graph
47	Low-Level Vision and Image Processing	Synergizing Global and Local Knowledge via Dynamic Focus Mechanism for Low-Light Image Enhancement
48	Low-Level Vision and Image Processing	Two-Stage Unsupervised Disentangled Realism Enhancement for Rendered Indoor Scene Images
49	Low-Level Vision and Image Processing	Corner Detection: Passive Non-Line-of-Sight Pedestrian Detection
50	Low-Level Vision and Image Processing	Label-Correlation Adaptive Central Similarity Hashing for Multi-Label Image Retrieval
51	3D Vision and Reconstruction	Visual Harmony: LLM's Power in Crafting Coherent Indoor Scenes from Images
52	3D Vision and Reconstruction	Superpixel Cost Volume Excitation for Stereo Matching
53	3D Vision and Reconstruction	Multi-view Depth Estimation with Adaptive Feature Extraction and Region-Aware Depth Prediction
54	3D Vision and Reconstruction	3D Data Augmentation for Driving Scenes on Camera
55	3D Vision and Reconstruction	A Pose-Aware Auto-Augmentation Framework for 3D Human Pose and Shape Estimation from Partial Point Clouds
56	3D Vision and Reconstruction	Efficient Emotional Talking Head Generation via Dynamic 3D Gaussian Rendering
57	3D Vision and Reconstruction	Generalizable Geometry-aware Human Radiance Modeling from Multi-view Images
58	3D Vision and Reconstruction	AG-NeRF: Attention-guided Neural Radiance Fields for Multi-height Large-scale Outdoor Scene Rendering
59	3D Vision and Reconstruction	JPA: A Joint-Part Attention for Mitigating Overfocusing on 3D Human Pose Estimation
60	3D Vision and Reconstruction	Realistic and Visually-pleasing 3D Generation of Indoor Scenes from a Single Image

61	3D Vision and Reconstruction	AttenPoint: Exploring Point Cloud Segmentation through Attention-Based Modules
62	3D Vision and Reconstruction	Multi-view 3D Reconstruction by Fusing Polarization Information
63	3D Vision and Reconstruction	Trajectory-based Calibration for Optical See-Through Head-Mounted Displays without Alignment
64	3D Vision and Reconstruction	Maximum Spanning Tree for 3D Point Cloud Registration
65	3D Vision and Reconstruction	Learning the Dynamic Spatio-Temporal Relationship Between Joints for 3D Human Pose Estimation
66	3D Vision and Reconstruction	MaskEditor: Instruct 3D Object Editing with Learned Masks
67	3D Vision and Reconstruction	DyGASR: Dynamic Generalized Gaussian Splatting with Surface Alignment for Accelerated 3D Mesh Reconstruction
68	3D Vision and Reconstruction	MMIDM: Generating 3D Gesture from Multimodal Inputs with Diffusion Models
69	3D Vision and Reconstruction	MagicGS: Combining 2D and 3D Priors for Effective 3D Content Generation
70	3D Vision and Reconstruction	ESD-Pose: Enhanced Semantic Discrimination for Generalizable 6D Pose Estimation
71	3D Vision and Reconstruction	Trans-DONeRF for Transparent Object Rendering with Mixed Depth Prior
72	3D Vision and Reconstruction	SFDNeRF: A Semantic Feature-Driven Few-Shot Neural Radiance Field Framework with Hybrid Regularization
73	3D Vision and Reconstruction	TriEn-Net: Non-parametric Representation Learning for Large-Scale Point Cloud Semantic Segmentation
74	3D Vision and Reconstruction	Decomposed Latent Diffusion Model for 3D Point Cloud Generation
75	3D Vision and Reconstruction	Learning Multi-Branch Attention Networks for 3D Face Reconstruction
76	3D Vision and Reconstruction	CP-VoteNet: Contrastive Prototypical VoteNet for Few-Shot Point Cloud Object Detection
77	3D Vision and Reconstruction	Cross Modality Fusion Network with Feature Alignment and Salient Object Exchange for Single Image 3D Shape Retrieval
78	3D Vision and Reconstruction	Enhanced Spatial Adaptive Fusion Network For Video Super-Resolution
79	3D Vision and Reconstruction	Multi-3D Occlusion Mask Learning for Flexible Occlusion Removal in Neural Radiance Fields
80	3D Vision and Reconstruction	Sketch-Based 3D Shape Retrieval via Cross-Modal Contrastive Learning and Difficulty-Aware Uncertainty Regularization

81	3D Vision and Reconstruction	Residual Hybrid Attention Enhanced Video Super-Resolution with Cross Convolution
82	3D Vision and Reconstruction	SDFReg: Learning Signed Distance Functions for Point Cloud Registration
83	3D Vision and Reconstruction	Unfolding Gradient Graph Regularization for Point Cloud Color Denoising
84	3D Vision and Reconstruction	ER-SFM: EFFICIENT AND ROBUST CLUSTER-BASED STRUCTURE FROM MOTION
85	3D Vision and Reconstruction	Multimodal Token Fusion and Optimization for 3D Human Mesh Reconstruction with Transformers
86	Vision Applications and Systems	Immersive 6DOF Roaming with Novel View Synthesis from Single Outdoor Panorama
87	Vision Applications and Systems	Delving Deeper into Clean Samples for Combating Noisy Labels
88	Vision Applications and Systems	Vision-Language Knowledge Exploration for Video Saliency Prediction
89	Vision Applications and Systems	Variational Capsules for Image Analysis and Synthesis
90	Vision Applications and Systems	An Avatar-based Intervention System for Children with Autism Spectrum Disorder
91	Vision Applications and Systems	VS-LLM: Visual-Semantic Depression Assessment based on LLM for Drawing Projection Test
92	Vision Applications and Systems	A Robust and Real-Time RGB-D SLAM Method with Dynamic Point Recognition and Depth Segmentation Optimization
93	Vision Applications and Systems	Edge-enhanced super-resolution reconstruction of rock CT images
94	Vision Applications and Systems	Multi-Prototype Co-Saliency Model for Plant Disease Detection
95	Vision Applications and Systems	Advancing Surveillance Video Clarity and Transmission: A Real-time Video Super-Resolution Model with Background Information Awareness
96	Vision Applications and Systems	A Novel Anti-rounding Image Steganography Method for Improved UNet++
97	Vision Applications and Systems	Apple Leaf Disease Segmentation in the Wild: A Multi-Task Collaborative Learning Approach
98	Vision Applications and Systems	EG-Trans: Transparent object segmentation with edge enhanced and global integrated Transformers
99	Vision Applications and Systems	CNN-Transformer with Stepped Distillation for Fine-Grained Visual Classification
100	Vision Applications and Systems	ALMRR: Anomaly Localization Mamba on Industrial Textured Surface with Feature Reconstruction and Refinement
101	Vision Applications and Systems	Adaptive Dual Attention Fusion Network for RGB-D Surface Defect Detection
102	Vision Applications and	Vehicle Appearance Dataset

	Systems	
103	Vision Applications and Systems	RefineStyle: Dynamic Convolution Refinement for StyleGAN
104	Vision Applications and Systems	Feature Refinement and Calibration for Continual Visual Search
105	Vision Applications and Systems	Joint Multi-Person Body Detection and Orientation Estimation via One Unified Embedding
106	Vision Applications and Systems	ULNet: A Lightweight Segmentation Network For Lane Detection
107	Remote Sensing Image Interpretation	Shape-Aware Soft Label Assignment and Context Enhancement for Oriented Object Detection
108	Remote Sensing Image Interpretation	Chareption: Change-Aware Adaption Empowers Large Language Model for Effective Remote Sensing Image Change Captioning
109	Remote Sensing Image Interpretation	Spectral-Spatial Multi-view Sparse Self-Representation for Hyperspectral Band Selection
110	Remote Sensing Image Interpretation	Adaptive Cross-spatial Sensing Network for Change Detection
111	Remote Sensing Image Interpretation	HANet: Hierarchical Attention Network for Remote Sensing Images Semantic Segmentation
112	Remote Sensing Image Interpretation	BiReNet: Bilateral Network with Feature Fusion and Edge Detection for Remote Sensing Images Road Extraction
113	Remote Sensing Image Interpretation	Latent Feature Representation-Based Low Rank Subspace Clustering for Hyperspectral Band Selection
114	Remote Sensing Image Interpretation	DICMNet: Dynamic Irregular Resnet with Multi-direction Channel Remapping for Remote Sensing Road Extraction
115	Remote Sensing Image Interpretation	Discriminative Representation-based Classifier for Few-shot Remote Sensing Classification
116	Remote Sensing Image Interpretation	Hyperspectral Image Change Detection via Cross-Sample Slot Attention and Dual Gated Feed-Forward Network
117	Remote Sensing Image Interpretation	Spectral Channel-weighting CAT for Hyperspectral image Classification
118	Remote Sensing Image Interpretation	BFRNet: Bimodal Fusion and Rectification Network for Remote Sensing Semantic Segmentation
119	Remote Sensing Image Interpretation	SFFAFormer: An Semantic Fusion and Feature Accumulation Approach for Change Detection on Remote Sensing Images
120	Remote Sensing Image Interpretation	A Novel Multi-scale Feature Fusion based Network for Hyperspectral and Multispectral Image Fusion
121	Remote Sensing Image Interpretation	A Sidelobe-Aware Semi-Deformable Convolutional Ship Detection Network for Synthetic Aperture Radar Imagery

122	Vision Problems in Robotics, Autonomous Driving	Edge Assisted Fast Optical Flow Matching SLAM in Underground Rescue Environments
123	Vision Problems in Robotics, Autonomous Driving	ViPro-BEV: Few-Shot Visual Prompting for Bird's-Eye-View Perception
124	Vision Problems in Robotics, Autonomous Driving	DNIV-SLAM: Neural Implicit Visual SLAM in Dynamic Environments
125	Vision Problems in Robotics, Autonomous Driving	BEVDot: Enhancing Environmental Perception for Autonomous Driving with a Deformable Depth Mechanism
126	Vision Problems in Robotics, Autonomous Driving	UDA-KB: Unsupervised domain adaptation RGB-Thermal semantic segmentation via knowledge bridge
127	Vision Problems in Robotics, Autonomous Driving	VPFNET : A SCALE-ADAPTIVE VOXEL POINT FUSION NETWORK FOR SEMANTIC SEGMENTATION OF POINT CLOUDS
128	Vision Problems in Robotics, Autonomous Driving	VTMF2N: Towards Accurate Visual-tactile Slip Detection via Multi-modal Feature Fusion in Robotic Grasping
129	Vision Problems in Robotics, Autonomous Driving	HHATP: A Lightweight Heterogeneous Hierarchical Attention Model for Trajectory Prediction
130	Vision Problems in Robotics, Autonomous Driving	Improved End-to-End Multilevel NeRF-Based Dense RGB-D SLAM
131	Vision Problems in Robotics, Autonomous Driving	KA-Seg: Improving LiDAR Point Cloud Segmentation through Key Point Sampling and Attention-based Querying
132	Vision Problems in Robotics, Autonomous Driving	ASPVNet: Attention Based Sparse Point-Voxel Network for 3D Object Detection
133	Vision Problems in Robotics, Autonomous Driving	Task-Oriented Scanpath Prediction with Spatial-Temporal Information in Driving Scenarios
134	Vision Problems in Robotics, Autonomous Driving	Intelligent Navigation System that Gives Trajectory Guidance in 3D Scenes
135	Vision Problems in Robotics, Autonomous Driving	Dynamic Attention-Enhanced Spatio-Temporal Network For Pedestrian Collision Risk Assessment
136	Vision Problems in Robotics, Autonomous Driving	Dynamic Object Suppression in Visual Odometry via Adaptive Masked Flow Refinement

海报 4		
序号	类别	论文题目
1	Neural Network and Deep Learning	Auto-USOD: Searching Topology for Underwater Salient Object Detection
2	Neural Network and Deep Learning	MBA-NER: Multi-Granularity Entity Boundary-Aware Contrastive Enhanced for Two-stage Few-Shot Named Entity Recognition
3	Neural Network and Deep Learning	Enhancing Zero-Shot Anomaly Detection: CLIP-SAM Collaboration with Cascaded Prompts

4	Neural Network and Deep Learning	Towards Adversarial-Robust Class-Incremental Learning via Progressively Volume-up Perturbation Generation
5	Neural Network and Deep Learning	Neighborhood Difference-Enhanced Graph Neural Network based on Hypergraph for Social Bot Detection
6	Neural Network and Deep Learning	SRMAE: Masked Image Modeling for Scale-Invariant Deep Representations
7	Neural Network and Deep Learning	An Entropy-based Pseudo-Label Mixup Method for Source-Free Domain Adaptation
8	Neural Network and Deep Learning	DAMS: Document Image Steganography with Dual Attention Multi-Scale Encoder-Decoder Architecture
9	Neural Network and Deep Learning	Dual-Task Cascaded for Proactive Deepfake Detection Using QPCET watermarking
10	Neural Network and Deep Learning	XrGroup: Graph Convolutional Networks for Group-Aware Pedestrian Trajectory Prediction with Speed information
11	Neural Network and Deep Learning	Invisible Backdoor Attack Through Singular Value Decomposition
12	Neural Network and Deep Learning	Self-supervised transformer-based pre-training method with General Plant Infection dataset
13	Neural Network and Deep Learning	Spatio-Temporal Perceiving Network Based Vision Transformer for 6-Hour Precipitation Prediction Using Multi-Meteorological Factors
14	Neural Network and Deep Learning	Learning Local Spatial and Global Context Activation for Visual Recognition
15	Neural Network and Deep Learning	CRFNet: A medical image segmentation method using the cross attention mechanism and refined feature fusion strategy
16	Neural Network and Deep Learning	SCC-CAM: Weakly Supervised Segmentation on Brain Tumor MRI with Similarity Constraint and Causality
17	Neural Network and Deep Learning	Global Structural Consistency Set Transformer
18	Neural Network and Deep Learning	IMO-Net: Integrated Memory Optimization Network for Video Instance Lane Detection
19	Neural Network and Deep Learning	Lightweight Facial Expression Recognition Based on Hybrid Multiscale and Multi-Head Collaborative Attention
20	Neural Network and Deep Learning	Single model learns multiple styles of Chinese calligraphy via Style Collection Mechanism
21	Neural Network and Deep Learning	FusionNet for Interactive Image Segmentation
22	Neural Network and Deep Learning	Dynamic Spatial-Temporal Perception Graph Convolutional Networks for Traffic Flow Forecasting

23	Neural Network and Deep Learning	Foreign object classification for coal conveyor belts based on deep learning
24	Neural Network and Deep Learning	Interpretable Unsupervised Homography Estimation
25	Neural Network and Deep Learning	DRC-NET: Density Reweighted Convolution Network for Edge Curve Extraction
26	Neural Network and Deep Learning	Unsupervised Underwater Image Enhancement Combining Imaging Restoration and Prompt Learning
27	Neural Network and Deep Learning	Generative Adversarial Imitation Learning Algorithm based on Improved Curiosity Module
28	Neural Network and Deep Learning	Zero-shot Blind Face Restoration via Conditional Diffusion Sampling
29	Neural Network and Deep Learning	Task-aware Few-shot Image Generation via Dynamic Local Distribution Estimation and Sampling
30	Neural Network and Deep Learning	Adversarial Training and Contrastive Learning with Bidirectional Transformers for Sequence Recommendation
31	Neural Network and Deep Learning	Empathizing Before Generation: A Double-layered Framework for Emotional Support LLM
32	Neural Network and Deep Learning	ST-RetNet: A Long-term Spatial-Temporal Traffic Flow Prediction Method
33	Neural Network and Deep Learning	DARTS-CGW: Research on Differentiable Neural Architecture Search Algorithm Based on Coarse Gradient Weighting
34	Neural Network and Deep Learning	PanoDthNet: Depth estimation based on indoor and outdoor panoramic images
35	Neural Network and Deep Learning	A Supervised Domain Adaptation Method with Alignment Regularization for Low-light Facial Expression Recognition
36	Neural Network and Deep Learning	DiffuSaliency: Synthesizing Multi-Object Images with Masks for Semantic Segmentation Using Diffusion and Saliency Detection
37	Neural Network and Deep Learning	EFOA: Enhancing Out-of-Distribution Detection by Fake Outlier Augmentation
38	Neural Network and Deep Learning	A Stereo Matching Method for Specular Objects via Cascaded Network and Joint Supervision
39	Neural Network and Deep Learning	An Asymmetric Game Theoretic Learning Model
40	Neural Network and Deep Learning	Learning 360° Optical Flow using Tangent Images and Transformer
41	Neural Network and Deep Learning	ODAdapter: An effective method of Semi-Supervised Object Detection for Aerial Images
42	Neural Network and Deep Learning	Frequency-domain Transformation-based Dynamic Gesture Recognition with skeleton

43	Neural Network and Deep Learning	MRGN: Multiscale Relation-gated Graph Network for Entity Alignment
44	Neural Network and Deep Learning	Adaptive Selective Knowledge Distillation: not blindly accepting teachers as Oracles
45	Neural Network and Deep Learning	Periodic Iterative Segmentation-Colorization Training: Line Drawing Colorization Using Text Tag with CBAMCat
46	Neural Network and Deep Learning	Histogram Prediction and Equalization for Indoor Monocular Depth Estimation
47	Neural Network and Deep Learning	SheepNet: Rapid Sheep Face Recognition Based on Attention and Knowledge Distillation
48	Neural Network and Deep Learning	LPMANet:A Lightweight Partial Multilayer Aggregation Network for Tiny Drone Detection
49	Neural Network and Deep Learning	HiTraj: Heterogeneous Interaction Learning with Transformers for Trajectory Prediction
50	Neural Network and Deep Learning	Adaptive Knowledge Matching for Exemplar-Free Class-Incremental Learning
51	Neural Network and Deep Learning	Focusing on Significant Guidance: Preliminary Knowledge Guided Distillation
52	Neural Network and Deep Learning	ESTOR:Enumerate-Specify-Tutor Mechanism Used of Lexicon in Chinese NER
53	Neural Network and Deep Learning	EBSD: Short Text Sentiment Classification Using Sentence Vector Enhancement Mechanism
54	Neural Network and Deep Learning	CEDP-YOLO: UAV Object Detection Based on Context Enhancement and Dynamic Perception
55	Neural Network and Deep Learning	TLLFusion: An End-to-End Transformer-Based Method for Low-Light Infrared and Visible Image Fusion
56	Neural Network and Deep Learning	BD-YOLO : High-precision lightweight concrete bubble detector based on YOLOv7
57	Neural Network and Deep Learning	Semantic Consistency-Enhanced Refined Hashing for Fine-Grained Image Retrieval
58	Neural Network and Deep Learning	Frequency Feature Enhanced Mix Calibration Attention Network for Sequential Recommendation
59	Neural Network and Deep Learning	CFMISA: Cross-modal Fusion of Modal Invariant and Specific Representations for Multimodal Sentiment Analysis
60	Neural Network and Deep Learning	A Privacy-Preserving Source Code Vulnerability Detection Method
61	Neural Network and Deep Learning	Physically Informed Prior and Cross-Correlation Constraint for Fine-grained Road Crack Segmentation
62	Neural Network and Deep Learning	AFSNet: Adaptive Feature Suppression Network for Remote Sensing Image Change Detection
63	Neural Network and Deep Learning	BIVL-Net: Bidirectional Vision-Language Guidance for Visual Question Answering

64	Neural Network and Deep Learning	Enhancing Task Identification through Pseudo-OOD Features for Class-Incremental Learning
65	Neural Network and Deep Learning	Contextual Feature-Based Medical Visual Question Answering Aided by Learnable Matrix
66	Neural Network and Deep Learning	ImgQuant: Towards adversarial defense with robust boundary via dual-image quantization
67	Neural Network and Deep Learning	Swelling-ViT: Rethink Data-efficient Vision Transformer from Locality
68	Neural Network and Deep Learning	Target-Specific Domain Adaptation via Geometry-Correlation Prediction for Point Cloud
69	Neural Network and Deep Learning	Dual-stream Network of Vision Mamba and CNN with Auto-scaling for Remote Sensing Image Segmentation
70	Neural Network and Deep Learning	A Novel Combined GAN for Defects Generation using Masking Mechanisms
71	Neural Network and Deep Learning	Semi-supervised lightweight fabric defect detection
72	Neural Network and Deep Learning	Semi-adaptive Synergetic Two-way Pseudoinverse Learning System
73	Neural Network and Deep Learning	Invariant Risk Minimization Augmentation for Graph Contrastive Learning
74	Neural Network and Deep Learning	Enhancing Fast Adversarial Training with Learnable Adversarial Perturbations
75	Neural Network and Deep Learning	DTAFORMER: Directional Time Attention Transformer For Long-Term Series Forecasting
76	Neural Network and Deep Learning	Unpaired Multi-scenario Sketch Synthesis via Texture Enhancement
77	Neural Network and Deep Learning	ISO-VTON: Fine-Grained Style-Local Flows with Dual Cross-Attention for Immersive Outfitting
78	Neural Network and Deep Learning	Near-surface Air Temperature Inversion Study Based on U-Net Family with Multi-source Data
79	Neural Network and Deep Learning	Relation Detection with Transformers for Panoptic Scene Graph Generation
80	Neural Network and Deep Learning	WEDNet: A Wavelet Enhanced Detail Network for Low-Light Image Enhancement
81	Neural Network and Deep Learning	Textureness-Aware Neural Network for Edge Detection
82	Neural Network and Deep Learning	Enhancing the Transferability and Stealth of Deepfake Detection Attacks Through Latent Diffusion Models
83	Neural Network and Deep Learning	Backdoor Richer Watermarks using Dynamic Mask Covering for Dual Identity Verification
84	Neural Network and Deep Learning	Pedestrian Trajectory Prediction using Spatio-Temporal VAE
85	Neural Network and Deep Learning	Real-Time Detection TRansformer with Bi-Level Routing Attention

86	Neural Network and Deep Learning	NFP-UNet: Deep Learning Estimation of Placeable Areas for 2D Irregular Packing
87	Neural Network and Deep Learning	Advancements in Photorealistic Style Translation with a Hybrid Generative Adversarial Network
88	Neural Network and Deep Learning	Transformer Image Quality Assessment Based on Multi-Directional Feature Extraction
89	Neural Network and Deep Learning	MORGAN: LightWeight Monaural Speech Enhancement using GAN Network
90	Neural Network and Deep Learning	Data augmentation guided Decouple Knowledge Distillation for low-resolution fine-grained image classification
91	Neural Network and Deep Learning	Virtual Student Distribution Knowledge Distillation for Long-tailed Recognition
92	Neural Network and Deep Learning	Open-Vocabulary Instance Segmentation-Boundary IS-Goal
93	Neural Network and Deep Learning	3DLaneFormer: End-to-End 3D Lane Detection with Voxel Descriptors
94	Neural Network and Deep Learning	More Like Real World Game Challenge for Partially Observable Multi-Agent Cooperation
95	Neural Network and Deep Learning	Centroid-centered Modeling for Efficient Vision Transformer Pre-training
96	Neural Network and Deep Learning	Spectral-Spatial Blockwise Masked Transformer With Contrastive Multi-View Learning for Hyperspectral Image Classification
97	Neural Network and Deep Learning	Local reactivation for communication efficient federated learning based on sparse gradient deviation
98	Machine Learning	Cluster center initialization for fuzzy K-modes clustering using outlier detection technique
99	Machine Learning	Generalizing soft actor-critic algorithms to discrete action spaces
100	Machine Learning	LarvSeg: Exploring Image Classification Data For Large Vocabulary Semantic Segmentation via Category-wise Attentive Classifier
101	Machine Learning	PhaseNN: An Unsupervised and Spatial-Frequency Integrated Network for Phase Retrieval
102	Machine Learning	Sequential Transfer of Pose and Texture for Pose Guided Person Image Generation
103	Machine Learning	Balanced Clustering with Discretely Weighted Pseudo-Label
104	Machine Learning	Tensor Robust Principal Component Analysis with Hankel Structure
105	Machine Learning	Self-Distillation via Intra-class Compactness
106	Machine Learning	An Enhanced Dual-Channel-Omni-Scale 1DCNN for Fault Diagnosis

107	Machine Learning	Visual-Guided Reasoning Path Generation for Visual Question Answering
108	Machine Learning	FedGC: Federated Learning on Non-IID Data via Learning from Good Clients
109	Machine Learning	Inter-class Correlation-based Online Knowledge Distillation
110	Machine Learning	Accelerating Domain Adaptation with Cascaded Adaptive Vision Transformer
111	Machine Learning	Multistage Compression Optimization Strategies for Accelerating Diffusion Models
112	Machine Learning	Defending Adversarial Patches via Joint Region Localizing and Inpainting
113	Machine Learning	Multi-view Spectral Clustering Based on Topological Manifold Learning
114	Machine Learning	Client selection mechanism for federated learning based on class imbalance
115	Machine Learning	A New Paradigm for Enhancing Ensemble Learning through Parameter Diversification
116	Machine Learning	Adaptive Multi-Information Feature Fusion MLP with Filter Enhancement for Sequential Recommendation
117	Machine Learning	FedDCP: Personalized Federated Learning Based on Dual Classifiers and Prototypes
118	Machine Learning	AtomTool: Empowering Large Language Models with Tool Utilization Skills
119	Machine Learning	Making the Primary Task Primary: Boosting Few-Shot Classification by Gradient-biased Multi-task Learning
120	Machine Learning	Cascade Large Language Model via In-Context Learning for Depression Detection on Chinese Social Media
121	Machine Learning	TRAE : Reversible Adversarial Example with Traceability
122	Machine Learning	A Two-stage Active Domain Adaptation Framework for Vehicle Re-Identification
123	Machine Learning	FBR-FL: Fair and Byzantine-Robust Federated Learning via SPD Manifold
124	Machine Learning	SecBFL-IoV: A Secure Blockchain-Enabled Federated Learning Framework for Resilience against Poisoning Attacks in Internet of Vehicles
125	Optimization and Learning Methods	Continuous Multi-Agent Path Finding for Drone Delivery
126	Optimization and Learning Methods	Enhancing Multi-modal Contrastive Learning via Optimal Transport-based Consistent Modality Alignment
127	Optimization and Learning Methods	Instance-level Scaling and Dynamic Margin-alignment Knowledge Distillation

128	Computational Photography, Sensing and Display Technology	Light Field Stitching via Mesh Deformation Alignment and Low-Rank-Based Fusion
129	Computational Photography, Sensing and Display Technology	CPE COIN++: Towards Optimized Implicit Neural Representation Compression via Chebyshev Positional Encoding
130	Computational Photography, Sensing and Display Technology	LF-SAET: Cascaded Spatial-Angular-EPI Transformers for Light Field Image Super-Resolution
131	Computational Photography, Sensing and Display Technology	A New Data-Driven Paradigm for SAR Jamming Suppression
132	Computational Photography, Sensing and Display Technology	Adaptive Decoupled Prompting for Class Incremental Learning

博士生论坛信息

论坛介绍: 论坛旨在为高校博士研究生提供一个学术交流和研讨平台,通过博士生学术报告、科研专家现场指导,大家交流学术思想、传递科研经验、增进彼此友谊,拓宽青年人才学术视野、培养创新能力、激发学术热情。

时间地点: 2024年10月20日17:30-21:00, 乌鲁木齐

论坛主席: 董晶(中科院自动化研究所), 左旺孟(哈工大)

交流嘉宾: 胡清华(天津大学), 程明明(南开大学), 魏云超(北交大), 高跃(清华大学), 姚鸿勋(哈尔滨工业大学), 邓成(西电)

论坛主席信息



论坛主席介绍: 董晶, 现任中国科学院自动化研究所多模态人工智能系统全国重点实验室研究员、博士生导师。主要从事计算机视觉、生物特征识别、多媒体内容取证与安全前沿方向的技术研究, 相关成果已成功应用于多个国家部委在多媒体智能鉴伪、生物特征大数据隐私保护等信息安全与国防安全方向的实际需求, 助力推动网络安全防护核心技术攻坚。入选国家高层次青年科技人才计划、先后以项目(课题)负责人承担4项国家自然科学基金、3项国家重点研发项目(课题)、国家重大专项子项及20余项省部级科研项目(课题)。在T-PAMI、TIFS、TCSVT等国际权威期刊及CVPR、AAAI、ACMM等学术会议上发表学术论文90余篇, 申请发明专利30余项, 其中已授权25项中国专利含3项美国专利。



论坛主席介绍: 左旺孟, 哈尔滨工业大学计算机学院教授、博士生导师。主要从事底层视觉、视觉生成、视觉理解和多模态学习等方面的研究。在CVPR/ICCV/ECCV/NeurIPS/ICLR等顶级会议和T-PAMI、IJCV及IEEE Trans.等期刊上发表论文200余篇, 谷歌学术引用60,000余次。曾任ICCV、CVPR等CCF-A类会议领域主席, 现任IEEE T-PAMI、T-IP、中国科学-信息科学、自动化学报等期刊编委。

交流嘉宾



交流嘉宾介绍：胡清华，国家优青/杰青，天津大学北洋讲席教授，天津市机器学习重点实验室主任，城市智能与数字治理教育部工程研究中心主任，CAAI 粒计算与知识发现专委会主任。从事大数据粒计算、多模态学习、不确定性建模和自主机器学习方面的研究，先后获得国家重点研发计划项目、国家自然科学基金重点项目、国家优青/杰青以及国防项目的资助。在 IEEE-TPAMI、IJCV、IEEE-TKDE、IEEE-TFS 等期刊以及 NeurIPS、CVPR、IJCAI、AAAI 等会议发表论文 300 余篇，获批发明专利 30 余个，出版专著 3 部，先后获得黑龙江省自然科学一等奖和天津市科技进步一等奖。目前担任 IEEE Trans. Fuzzy Systems，自动化学报、电子学报、智能系统学报等期刊的编委。



交流嘉宾介绍：程明明，南开大学杰出教授，主持承担了国家杰出青年科学基金、优秀青年科学基金项目、科技部重大项目课题等。他的主要研究方向是计算机视觉和计算机图形学，在 SCI 一区/CCF A 类刊物上发表学术论文 100 余篇(含 IEEE TPAMI 论文 30 余篇)，h-index 为 80，论文谷歌引用 4 万余次，单篇最高引用 4700 余次，多次入选全球高被引科学家和中国高被引学者。技术成果被应用于华为、国家减灾中心等多个单位的旗舰产品。获得教育部自然科学一等奖 2 项、其他省部级科技奖 2 项。培养的 3 名博士生获得省部级优秀博士论文奖。现担任中国图象图形学学会副秘书长、天津市人工智能学会副理事长和顶级期刊 IEEE TPAMI, IEEE TIP 和《中国科学：信息科学》编委。



交流嘉宾介绍：魏云超，北京交通大学教授，国家高层次人才计划获得者。曾在新加坡国立大学、美国伊利诺伊大学厄巴纳-香槟分校、悉尼科技大学从事研究工作。主要研究方向包括面向非完美数据的视觉感知、多模态数据分析、生成式人工智能等。入选 MIT TR35 China、百度全球高潜力华人青年学者、《澳大利亚人》TOP 40 Rising star，获世界互联网大会领先科技奖(2023)、教育部高等学校自然科学一等奖(2022)、中国图象图形学学会科技技术奖一等奖(2019)、澳大利亚研究委员会青年研究奖(2019)、IBM C35R 最佳研究奖(2019)、计算机视觉世界杯 ImageNet 目标检测冠军(2014)及多项 CVPR 竞赛冠军。



交流嘉宾介绍：高跃，清华大学长聘副教授、博士生导师，国家青年特聘专家、DeepTech 2022 年中国智能计算科技创新人物。主要研究领域为人工智能、计算机视觉及医学图像处理，提出并发展了超图计算理论及方法，实现了高阶关联的结构建模与计算，有效应对数据不足和信息关联复杂等挑战，应用于视觉目标检测、医学辅助诊断等领域。研究成果在 IEEE TPAMI 等发表论文百余篇，论文引用万余次，出版《Hypergraph Computation》等英文专著，多次入选科睿唯安高被引科学家和爱思唯尔中国高被引学者，担任 International Journal of Computer Vision 和 Medical Image Analysis 等国际重要期刊编委。曾获得中国电子学会自然科学一等奖、福建省科技进步一等奖等。

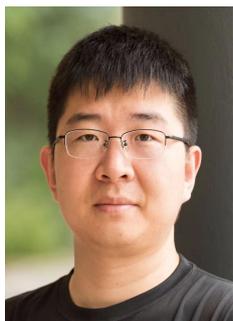


交流嘉宾介绍：姚鸿勋，哈尔滨工业大学长聘教授，黑龙江省政府特殊津贴专家，教育部“新世纪优秀人才”，AI 全球 2000 位最具影响力学者 (AMiner 机构发布)，哈工大人工智能专业负责人。任中国图象图形学学会常务理事，中国图象图形学学会情感计算与理解专业委员会主任，黑龙江省人工智能学会常务理事，中国计算机学会多媒体专委会委员；主要研究领域为计算机视觉智能、多媒体数据分析与理解、模式识别、情感计算。发表 ICCV, CVPR, ACM MM 等顶级国际会议及 TPAMI, IJCV, TIP, TMM 等高影响因子国际期刊文章学术论文 200 余篇，H 指数>40。



交流嘉宾介绍：邓成，西安电子科技大学教授、博士生导师，分别在西安电子科技大学取得学士、硕士和博士学位。2012 年入选教育部新世纪优秀人才，2017 年入选陕西省中青年科技创新领军人才，2018 年获得陕西省青年科技奖。2020 年入选长江学者特聘教授、国家百千万人才工程、陕西省师德标兵。主要研究多模态数据智能分析与推理。主持包括国家自然科学基金面上基金、科技部“863”计划、陕西省重点研发计划等科研课题近 30 项。近 5 年，在国际 SCI 一区/二区刊物 IEEE T-NNLS、T-CYB、T-IP、T-MM、T-CSVT 等发表论文 40 余篇，在 CCF A 类会议 ICML、NeurIPS、ICCV、CVPR、KDD、AAAI、IJCAI 等发表论文 50 余篇。现担任国际知名期刊 Pattern Recognition、Neurocomputing、Pattern Recognition Letters 的副编辑；担任多个国际学术会议的领域主席、高级程序委员等，如 CVPR2021、IJCAI2020、IJCAI2019、ICME2020、ICPR2018 等，以及近 20 余个国际刊物的审稿人，如 IEEE T-PAMI、T-IP、T-CYB、T-NNLS、IJCV 等。

讲者



博士生讲者简介：常东良，清华大学自动化系博士后，国家资助博士后研究人员计划（B 档）获得者，研究方向为细粒度图像理解。共发表论文 40 余篇，其中 15 篇为 CCF A 类论文，谷歌学术引用超 2000 次。主持国家青年基金和中国博士后科学基金面上项目。

报告题目：基于信息融合的细粒度图像分类方法研究

摘要报告：细粒度图像分类是指对属于同一父类的子类进行识别，比如区分黑脚信天翁和灰背信天翁，其难点在于较大的类内差异和较小的类间相似性。近年来，细粒度图像分类算法已经能够在超过 17 种细粒度数据库上获得超过 90% 的性能，具备了推向现实场景的能力。然而，我们通过大规模的 human study 实验发现，大多数现有的细粒度视觉分类模型仅输出单一粒度标签的方式存在着严重的实用性问题，不能满足所有人的需求--具有不同专业/知识背景的人，想要的答案是不一样的。为此，我们提出了基于解耦增强网络的多粒度图像分类算法，实现了跨粒度的信息融合。此外，我们提出了基于跨模态信息融合的细粒度图像分类算法，尝试如何将 AI 模型掌握的细粒度知识传递给普通人，让你我这些（鸟类）识别的外行，在知识的帮助下，也能够看出这些（鸟类）的差异，进而识别他们。本次报告将以上述两个工作为主体，向大家汇报我博士阶段的工作进展。



博士生讲者简介：陈翔，南京理工大学博士生，师从潘金山教授。研究方向为恶劣环境下的视觉增强。目前在 CCF/CAAI-A 类会议和期刊上发表多篇论文。相关研究工作获华为终端 Camera 学术之星、首届“中国宇航学会高水平学位论文激励计划”等荣誉奖励。

报告题目：恶劣环境下的视觉增强与计算

摘要报告：图像视频作为计算机感知场景的关键，其质量在很大程度上决定了智能感知算法的鲁棒性。然而复杂环境下图像视频获取的过程容易受恶劣天气环境的影响，导致所获取的图像视频质量低、内容看不清。如何实现恶劣天气成像环境中的图像视频清晰化从而提升现有方法对图像视频内容理解的性能是解决恶劣环境下自适应感知的关键。本次报告以雨天场景为例，介绍恶劣环境下的视觉增强与计算相关研究工作。



博士生讲者简介：段沛奇，于 2023 年从北京大学计算机学院获得理学博士学位。现为北京大学计算机学院博雅博士后，从事神经形态相机辅助高质量成像研究，在相关领域发表论文 11 篇。入选 2023 年度“博士后创新人才支持计划”（“博新计划”），主持国家自然科学基金青年科学基金、博士后科学基金面上资助等多个项目。

担任 IEEE TPAMI, TIP, TMM, TCSVT, ACM 等重要国际期刊审稿人,担任 CVPR, ICCV, ECCV, NeurIPS, AAAI, ICLR 等重要国际学术会议审稿人。

报告题目：神经形态事件与传统图像的互补增强机制

摘要报告：脉冲相机、事件相机等神经形态相机，对比逐帧成像的普通相机，拥有独特的优势，尤其是其对于高速运动物体和高动态范围场景的感知能力。本报告以事件相机为例，介绍神经形态与传统相机融合并互补增强成像质量和提升场景感知的计算摄像方法与系统，具体从三个方面展开：首先，介绍传统图像如何增强事件信号的质量，对其进行去噪、超分辨率操作并建立其与传统图像的关键点数据匹配；其次，介绍事件信号如何引导提升图像动态范围，消除图像模糊，矫正卷帘快门果冻效应。最后，介绍事件信号如何结合图像实现快速的直接/全局光照分解。



博士生讲者简介：顾峥，南京大学计算机学院在读博士，导师为高阳教授，同时他是香港城市大学电脑科学系联合培养博士，导师为廖菁副教授，他的研究方向为开放场景下的视觉感知与生成。

报告题目：基于扩散模型的视觉上下文学习

摘要报告：上下文学习（In-Context Learning, ICL）作为一种新的学习范式，能够通过大模型的类比推理能力在有限的示例对输入下完成各类生成任务。然而，与语言模型相比，视觉上下文学习在任务理解能力、学习效率与任务泛化性等方面仍存在诸多挑战。本报告回顾视觉上下文学习的发展脉络，并介绍一种基于预训练扩散模型的视觉上下文学习方法，该方法能够同时利用视觉提示与文本提示，通过类比推理的方式执行各类视觉生成任务，最后探讨目前存在的问题以及未来发展方向。



博士生讲者简介：黄文柯，武汉大学计算机学院人工智能系 23 级博士生，导师为叶茫教授和杜博教授。研究方向为联邦学习。目前已于 CCF-A 类会议和期刊上发文 17 篇，第一和共一 14 篇，包含 IEEE TPAMI 2 篇、CVPR 3 篇。曾在微软亚洲研究院、阿里巴巴集团担任研究实习生，主持国家自然科学基金青年学生基础研究项目。

报告题目：异构联邦学习

摘要报告：联邦学习作为一种安全隐私多方合作范式，能够实现数据可用不可见，数据不动模型动。因而，在数据隐私受限的场景下，联邦学习可以赋能数据敏感需要大量数据的领域，譬如大模型训练，医疗等领域。而现有联邦学习体系通常建立在数据相同分布且模型相同架构的设定，而实际环境下存在如下问题：数据采集方式不同数据分布差异大、参与者合作意图难辨别恶意攻击抵挡难、用户自行设计模型模型结构不相同。针对以上问题，提出面向多域泛化的联邦领域约束学习方式、异构场景下恶意行为的统一建模与解耦合联邦框架等理论研究方法，有效推动联邦学习在实际部署应用场景下的泛化性，鲁棒性和兼容性，也为面向安全可靠和自主可控的新一代人工智能研究及实际应用提供重要的技术支撑和理论。



博士生讲者简介：寇人可（1992.11），男，陕西西安人，工程师，陆军工程大学石家庄校区博士研究生。他的研究兴趣包括计算机视觉和深度学习，重点是红外小目标检测与跟踪算法的设计与嵌入式部署。他在攻读博士学位期间，以第一作者发表学术论文 10 篇，其中 SCI 6 篇（含 2 篇 ESI 高被引论文，5 篇中科院 2 区以上 Top 期刊论文），合作发表论文 10 篇；申请国家发明专利 3 项（第一发明人）；荣获空军首届航空创意挑战赛二等奖 2 项（分别为第 1，2 完成人）；荣获第三届中国图象图形学报研究生学术论坛杰出报告一等奖；主持项目 1 项，参与项目 2 项；撰写专著 1 部（待出版）；担任 IEEE-TGRS、IEEE-JSTARS、MDPI、Senors 等多个期刊的审稿人。

报告题目：红外小目标检测技术进展与开源测评框架探索

摘要报告：远距离红外小目标检测技术被广泛运用在侦察、预警等领域，其检测性能和效率将直接决定行动的成功与否，任何微小的延时和误判都可能导致严重的后果。因此，如何提升红外小目标检测技术的性能，确保准确、迅速识别并跟踪远距离目标一直是计算机视觉领域研究的热点。近几十年来有大量的红外小目标检测算法被相继提出，这些算法经历了从低层特征建模到高层深度学习的巨大跃进，而与之伴随的评价指标呈现出多样化和不统一的特点。为进一步规范该领域评价体系，此次报告旨在系统梳理红外小目标检测领域的评价指标，并在深入调研的基础上提出客观统一的评价标准。具体而言，首先以红外小目标检测技术的发展历程为脉络回顾了与之伴随的评价指标的演进过程；其次，全面总结了目前广泛使用的评价指标，剖析了其优缺点和内在联系。为进一步了解领域专家的观点和使用习惯，我们对 39 位知名学者进行了问卷调查，并统计分析了 91 篇权威文献，重点探求了评价指标细节设置上的异同。综合调研结果，给出了优化的评价指标体系和配套的使用规范，并对现有评价指标的局限性进行了讨论，提出了未来改进和拓展的研究方向。



博士生讲者简介: 李焜东, 主要研究方向包括自然语言处理、知识图谱和多模态机器学习; TencentPretrain 预训练框架和 Linly 大模型项目主要作者, GitHub Star 数量超 5000; 中文模型测评社区 CLUEbenchmark 创始成员; 从头构建千亿级别通用/领域大模型。在 AI 领域发表多篇 CCF-A 类论文, 多次国际竞赛获奖。

报告题目: 人脸图文多模态生成式预训练模型

摘要报告: 报告简述了人脸领域图文多模态预训练工作, 包括人脸双向图文生成模型、基于自监督的人脸编辑模型、大规模人脸领域图文数据集及对比学习模型等。

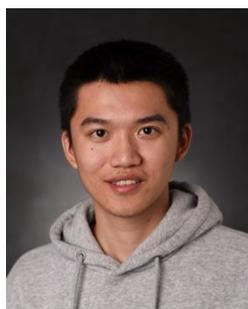


博士生讲者简介: 李泽宇, 北京邮电大学人工智能学院博士二年级在读, 指导老师为张曼教授, 研究方向为 3D 场景重建、3D 资产表面材质生成。

报告题目: 基于 2D 先验的 3D 资产材质生成

摘要报告: 在强大的 2D 生成先验的驱动下, 最近的研究已经成功地从文本或视觉引导中自动创建 3D 对象。然而, 这种二维图像先验会影响光照和阴影的效果。使用分数蒸馏采样(SDS)优化的材料图不可避免地涉及无关信息。缺乏精确的材料使得生成的资产在新场景中难以重新渲染, 这限制了它们在下游应用中的可用性。

从艺术家的角度, 可以根据对象的外观和语义推断出材料信息。受此启发, 我们提出一个 3D 资产材料生成框架 MaterialSeg3D, 首次采用基于感知的分割方法, 直接从物体的多视图渲染中预测材料, 并有 2D 先验的监督。



博士生讲者简介: 梁科在知识表示与推理, 图机器学习等方面提出多个原创算法, 以第一/共一完成 CCF A 或 Trans. 论文 10 余篇, 参与 10 余项国家、省部级项目; 参与 CAAI 青工委, 并参与 TKDE, ICLR, KDD 等 35 个期刊会议审稿工作, 获国防大年度优秀学员等奖项。

报告题目: 结构化知识表示与推理

摘要报告: 现有结构化知识表示普遍面临图结构先验信息利用少、全图信息聚合慢、多源信息融合难等问题, 影响了方法的推理能力。因此, 如何设计强表达性、高效率和多源性的结构化知识表示学习方法是知识图谱相关研究中亟待开展的课题。汇报者在博士期间致力于克服现有结构化知识表示与推理中存在的研究难点, 从静态知识图谱和引入额外信息的时序、多模态知识图谱两个角度出发, 探索了现有包括对比学习、图神经网络、Transformer、多模态学习在内的多种先进人工智能技术与知识图谱表示学习方法相结合的可行方案, 以提升现有方法的表征质量, 为推动该领域的发展提供新思路、新技术和新途径。



博士生讲者简介：刘猛，国防科技大学博士生，师从刘新旺教授。方向为图神经网络和聚类。一作在 ICLR、MM、TNNLS 等刊物发表论文 9 篇，被引 500 余次。获 2024 中国算力大会最佳论文，CCHI2023 最佳学生论文、国家奖学金等。担任 TKDE、TNNLS、NeurIPS、ICML 等审稿人。

报告题目：深度时序图聚类

摘要报告：深度图聚类最近受到了广泛的关注，因其能够增强模型在无监督场景下的表示学习能力。然而，作为可以捕获关键的动态交互信息的时序图，其上的深度聚类尚未得到充分的探索。这意味着在许多面向聚类的现实场景中，时序图只能作为静态图来处理。这不仅会导致动态信息的丢失，还会引发巨大的计算消耗。为了解决这个问题，本文提出了一种称为 TGC 的深度时序图聚类框架，通过改进深度聚类技术来适应基于交互序列的时序图批处理模式。此外，本文从多个层面讨论了时间图聚类与静态图聚类的区别。大量在 TGC 框架上的实验表明，时序图聚类可以更灵活地在时间和空间需求之间找到平衡。



博士生讲者简介：马梅，宁夏大学物理学博士研究生，宁夏大学 HCTC 团队核心成员。曾荣获省级优秀硕士学位论文。主要研究方向：医学图像生成、材料智能计算。

报告题目：基于图第一性原理的模式生成

摘要报告：模式生成是 AIGC 领域中的核心重要任务。目标模式作为模式生成的重要组成部分，其符合物理真实和理论一致特性为模式生成提供更高的可靠性和有效性。针对符合特定约束的目标模式生成技术通常存在着一致性差、生成质量难以控制等问题。基于图第一性原理的模式生成结合图论与系统的起源性要素捕捉模式的本质特征，并通过第一性原理确保生成模式符合物理真实和理论一致性，实现对复杂模式的生成与构建。具体研究内容包括：1) 基于等势图的跨模态生成方法，旨在挖掘不同模态之间的内在联系，提升生成模型在多样性和一致性上的表现；2) 基于孪生图的多标签生成方法，构建彼此关联的孪生图结构，有效促进多标签数据的生成与优化，解决传统方法中因标签稀疏性带来的生成挑战；3) 基于等变图的新材料生成方法，聚焦于材料特性与结构对称性之间的内在联系，利用等变性原理提升新材料的设计效率和性能。



博士生讲者简介：马彦彪，西安电子科技大学人工智能学院 2020 级直博生，师从焦李成教授。以第一作者在 IJCV, ICLR, CVPR, ACM MM, TMM, JSTARS 等期刊和会议发表学术论文 7 篇。获得由 CVPR, ICCV, ECCV 以及 IGARSS 组织的国际竞赛冠军 5 项，亚军和季军各 1 项，其中包括多时相，多模态，多光谱数据用于无电居民区检测竞赛冠军，洪水灾害中的无人机航拍图像视觉问答竞赛冠军等。长期担任 AAAI, NeurIPS, ICML, ICLR, CVPR, TNNLS, TMM 等会议和期刊审稿人。

报告题目：通用场景下深度神经网络不公平问题的几何机制与优化

摘要报告：样本不平衡被广泛认为是模型偏差的来源，为了解决数据长尾分布的挑战，研究人员提出了多种方法来减少模型偏差，其中大多数方法假设样本较少的类别是弱类别。然而，最近的研究表明，即使在完全平衡的数据集上进行训练，深度神经网络也会表现出类别偏见，这是现有理论无法解释的现象。为了解决和改善深度神经网络中更广泛的不公平问题，系统的提出了模型公平性的几何分析视角，提出了深度神经网络中感知流形的概念,和一系列用于深度神经网络中感知流形几何特性的度量。通过一系列实验发现，感知流形的几何特性与模型公平性之间存在显著相关性。一系列工作从几何角度解释了模型偏差的产生机制，提醒研究人员关注非长尾甚至样本均衡数据集上的模型偏差问题。



博士生讲者简介：汪茜茜，安徽大学计算机科学与技术学院新近进站博后，2024 年获得安大工学博士学位，研究兴趣包括结构化建模，视觉表示学习，小样本学习等。主要研究成果发表在 IJCV, IEEE T-CSVT, IEEE T-MM, Physics in Medicine and Biology 等期刊，现为 T-CSVT, PR 等期刊审稿人。

报告题目：基于结构化 Transformer 模型的视觉表示学习

摘要报告：视觉表示学习（Visual Representation Learning）是计算机视觉领域中的一项核心任务，旨在从原始的像素数据中提炼出能够揭示图像本质属性和语义信息的有效特征，从而为图像分类、目标检测、图像检索等后续任务提供坚实基础。随着深度学习技术的飞速发展，特别是 Transformer 模型的引入，通过将 Transformer 与卷积神经网络（CNN）相结合或完全替代 CNN，可以实现对图像数据更为高效的表示学习。然而，现有的视觉 Transformer 方法大多沿用传统的 Transformer 架构，这存在以下问题。首先，这些方法主要关注图像内上下文关系建模，忽视了不同图像样本间的相关性；其次，这些方法在处理多模态数据时忽视了不同模态图像间固有结构关系建模，使得模型可能因模态差异问题而获得次优性能。本报告将介绍我们在这两个问题上的相关研究，旨在为视觉表示学习带来新的方法和思路。



博士生讲者简介：Fengyi Wu is pursuing his Ph.D. at the School of Information and Communication Engineering, University of Electronic Science and Technology of China (UESTC), supervised by Prof. Zhenming Peng. Currently, his research interests focus on optimization-based methods with applications in signal and image processing, learning, and computer vision.

报告题目：基于模型驱动的红外小目标检测网络

摘要报告：随着深度学习的兴起与数据集的丰富，基于神经网络的红外小目标检测得到了长足的长进。然而，大多数方法专注于利用主流网络对红外图像进行处理，忽略了红外信息的本质，导致了一定的“黑盒”效应。为了解决以上问题，我们专注将传统可信的红外小目标方法进行深度展开，将具有物理信息的传统算子及低秩稀疏等传统算法进行网络化实现，打造一系列近“白盒”的红外小目标检测网络。



博士生讲者简介：杨杰 香港中文大学深圳

报告题目：以人为中心的感知，理解，推理和生成

摘要报告：在动态复杂的环境中，构建一个能够感知、分析和解释人类与物体状态及其潜在动机的系统，依然是一个重大且充满挑战的问题。本报告围绕开放世界下的以人为中心的感知、理解、推理和生成，提出了三个核心研究问题：（1）X-Pose: Detecting Any Keypoints（专注于感知任意物体）；（2）Open-World Human-Object Interaction Detection via Multi-modal Prompts（旨在感知人和物体的交互状态）；（3）F-HOI: Toward Fine-grained Semantic-Aligned 3D Human-Object Interactions（探索三维环境中人-物交互关系的理解，推理与生成）



博士生讲者简介：周尚辰：新加坡南洋理工大学 MMLab@NTU 博士生，师从 Chen Change Loy 教授。他在计算机视觉领域的顶级会议和期刊，如 CVPR、ICCV、ECCV、NeurIPS、TPAMI 等上发表了多篇论文，谷歌引用超过 3600 次。他曾获得国家优秀自费留学生奖学金，WAIC 青年优秀论文提名奖，ICIMCS 最佳论文奖，并在 NTIRE 视频复原比赛中获得三项冠军。他还与 ECCV2022、CVPR2023 和 CVPR2024 会议上参与组织了移动智能摄影和成像（MIPI）国际系列研讨会。目前主要研究兴趣包括图像/视频增强、

生成模型。

报告题目: Unleashing Diffusion Models for Low-Level Vision

摘要报告: 最近,随着视觉生成模型的飞速发展,它们正逐渐成为探索底层视觉任务的强大引擎。生成模型凭借其卓越的生成能力,显著提升了底层视觉模型的性能,使得输出画质有了质的提升。然而,这种强大的生成能力也如同一把双刃剑,模型输出常常伴随着纹理改变,给底层视觉模型的设计带来新的问题和挑战。本报告将探讨 Diffusion 生成模型如何驱动不同底层视觉任务的探索与进步。



博士生讲者简介: 朱明丽,香港中文大学(深圳)数据科学学院在读博士生,师从吴保元教授,主要研究方向为可信人工智能,研究后门学习领域的攻击和防御问题。在 ICCV、NeurIPS、CVPR 等国际顶级会议发表论文四篇。

报告题目: Breaking the False Sense of Security in Backdoor Defense through Re-Activation Attack

摘要报告: 神经网络在防御后门攻击方面一直面临挑战,导致攻防双方持续较量。虽然现有的后门防御策略在降低攻击成功率方面表现出色,但我们能否自信地宣称后门威胁已真正从模型中消除?为此,我们重新研究了防御后被植入后门的模型(称为防御模型)的特征。令人惊讶的是,我们发现在现有的训练后防御策略衍生的防御模型中,原始后门仍然存在,后门的存在性通过一个称为后门存在系数的新指标来衡量。这意味着后门只是处于休眠状态,而不是被消除。为了进一步验证这一发现,我们通过实证表明,这些休眠的后门可以在推理过程中轻松地重新激活,方法是使用通用对抗攻击通过精心设计的微小扰动来操纵原始触发器。更实际的是,我们将后门重新激活扩展到黑盒场景,其中防御模型只能在推理期间被对手查询,并开发两种有效的方法,即基于查询和基于传输的后门重新激活攻击。在图像分类和多模态对比学习(即 CLIP)任务上验证了所提出方法的有效性。总之,这项工作揭示了现有防御策略中从未探索过的关键漏洞,强调了未来设计更强大、更先进的后门防御机制的紧迫性。

竞赛论坛

- 17:30-17:35 开幕致辞
- 17:35-17:45 赛道 1 主办方报告及颁奖：大规模合成孔径雷达图像多类别目标检测竞赛
- 17:45-17:55 赛道 1 获奖队伍报告
- 17:55-18:05 赛道 2 主办方报告：广域红外小目标检测挑战赛
- 18:05-18:15 赛道 2 获奖队伍报告
- 18:15-18:25 赛道 3 主办方报告：大规模 X 光安检图像检测挑战赛
- 18:25-18:35 赛道 3 获奖队伍报告
- 18:35-18:45 赛道 4 主办方报告：行车场景下认知与决策
- 18:45-18:55 赛道 4 获奖队伍报告
- 18:55-19:05 赛道 5 主办方报告：基于多模态大模型的智慧城市治理图像识别挑战赛
- 19:05-19:15 赛道 5 获奖队伍报告
- 19:15-19:20 闭幕致辞

城市概括

乌鲁木齐（Urumqi），通称“乌市”，旧称迪化，新疆维吾尔自治区辖地级市、首府、I型大城市，位于中国西北部，是丝绸之路经济带核心区新疆维吾尔自治区的省会城市，也是世界上离海最远的城市，有“亚心之都”的称呼。乌鲁木齐是新疆的政治、经济、文化、科教和交通中心，国务院批复确定的中国西部地区重要的中心城市和面向中亚的国际商贸中心，是第二座亚欧大陆桥中国西部桥头堡和中国向西开放的重要门户。乌鲁木齐地形以山地、盆地、平原和丘陵为主，属于中温带大陆性干旱气候，具有明显的季节差异和较大的昼夜温差。截止2023年，全市辖七区一县，总面积13800平方公里，常住人口408.24万人，城镇化率96.56%，居住着汉、维吾尔、哈萨克、柯尔克孜、回、锡伯等52个民族。

18世纪中叶，清朝政府开始在乌鲁木齐驻军屯田，并鼓励移民实边，促进了当地的经济发展和人口增长。19世纪末，乌鲁木齐成为新疆的政治、经济和文化中心。1912年，中华民国成立后，乌鲁木齐继续保持着重要地位。20世纪30年代，新疆省政府在乌鲁木齐设立，进一步推动了城市的发展。抗日战争期间，乌鲁木齐成为中国西北地区的重要后方基地，接纳了大量的难民和流亡学生。同时，新疆与苏联的贸易往来增加，促进了乌鲁木齐的经济繁荣。1949年，新中国成立后，新疆和平解放，乌鲁木齐迎来了新的发展机遇。在国家的支持下，乌鲁木齐进行了大规模的基础设施建设和工业化进程，城市面貌发生了巨大变化。

改革开放以来，乌鲁木齐市经济快速发展，逐渐成为中国西部地区的重要城市。近年来，乌鲁木齐市在城市建设、交通、教育、文化等方面取得了显著成就。从20世纪开始大量的具有新疆地域民族风格的建筑不断涌现。乌鲁木齐拥有丰富的人文地理和美食资源，是一个充满活力和魅力的旅游目的地。乌鲁木齐是一个融合了多元文化、风景壮丽、资源丰富的现代化城市。其中天山天池：是一个高山湖泊，位于天山山脉的中部，湖水清澈碧绿，周围被群山环绕，景色壮观。红山公园：是一座以红山文化为主题的公园，园内有許多红山文化的遗址和文物，让你感受到古老文化的魅力。新疆国际大巴扎：是一个闻名世界的景点，这里有许多美味的特色菜肴，如烤羊肉串、手抓饭、大盘鸡等，这些美食不仅味道鲜美而且富有浓郁的地方特色，让味蕾得到极致的享受。乌鲁木齐博物馆：被誉为“地下文物宝库”，这里的文物见证了新疆的历史和文化。

乌鲁木齐是丝绸之路经济带的金融中心，中亚地区的区域商贸中心，中国辐射全疆及中亚的区域金融中心，围绕新疆油气生产加工、煤炭煤电煤化工、绿色矿业、粮油、棉花和纺织服装、绿色有机果蔬、优质畜产品、新能源新材料“八大产业集群”，经济以第三产业为支柱，已形成三二一的经济结构。其中，第一产业以种植业和畜牧业为主，利用设施农业和高新科技种植粮食、果蔬和经济作物；第二产业以石油化工、先进装备制造、新材料、纺织服装、新能源等新兴集群产业为重点，延长产业链，形成产业集群，向新型制造业转型；第三产业以电子商务、对外贸易、交通运输业和旅游业为主，借助区位优势、地方特色综合发展。中国新疆国际煤炭工业博览会是新疆维吾尔自治区的重要活动、亚洲两大煤炭展之一。

2023年，乌鲁木齐实现地区生产总值（GDP）4168.46亿元，按不变价计算，比上

年增长 6.0%。其中：第一产业增加值 33.12 亿元，增长 13.7%；第二产业增加值 1146.74 亿元，增长 0.9%；第三产业增加值 2988.60 亿元，增长 7.8%。三次产业结构为 0.8：27.5：71.7。人均地区生产总值 102078 元。



